

VISIT...

LANZAROTE
Caliente.COM

KIRK-OTHMER

ENCYCLOPEDIA OF

CHEMICAL

TECHNOLOGY

EDITING - by **WATCHER**

VOLUME 9

FOURTH EDITION

**Elastomers, Synthetic to
Expert Systems**

CONTENT INDEX (vol 9)

(with hyperlinks)

Edited by <http://www.watcherworld.narod.ru/>

(ELASTOMERS, SYNTHETIC)	
- POLYISOPRENE	2
- THERMOPLASTIC ELASTOMERS	11
ELECTRICAL CONNECTORS	23
ELECTRICALLY CONDUCTIVE POLYMERS	35
ELECTROANALYTICAL TECHNIQUES	49
ELECTROCHEMICAL PROCESSING	61
- Introduction	62
- INORGANIC	69
- ORGANIC	86
ELECTROLESS PLATING	106
ELECTRONIC MATERIALS	115
ELECTRONICS, COATINGS	121
ELECTROPHOTOGRAPHY	128
ELECTROPLATING	143
ELECTROSEPARATIONS	171
- ELECTRODIALYSIS	172
- ELECTROPHORESIS	178
EMBEDDING	187
EMULSIONS	196
ENAMELS, PORCELAIN OR VITREOUS	207
ENERGY MANAGEMENT	220
ENGINEERING, CHEMICAL DATA CORRELATION	232
ENGINEERING PLASTICS	261
ENVIRONMENTAL IMPACT	281
ENZYME APPLICATIONS	283
- INDUSTRIAL	284
- THERAPEUTIC	307
ENZYME INHIBITORS	318
ENZYMES IN ORGANIC SYNTHESIS	331
EPINEPHRINE AND NOREPINEPHRINE	354
EPOXY RESINS	362
ESTERIFICATION	374
ESTERS, ORGANIC	386
ETHANOL	401
ETHERS	423
ETHYLENE	431
ETHYLENE OXIDE	450
EVAPORATION	471
EXHAUST CONTROL, AUTOMOTIVE	480
EXHAUST CONTROL, INDUSTRIAL	500
EXPECTORANTS, ANTITUSSIVES, AND RELATED AGENTS	517
EXPERT SYSTEMS	530

ELASTOMERS, SYNTHETIC

Survey, Vol. 8,
Acrylic elastomers, Vol. 8,
Butyl rubber, Vol. 8,
Chlorosulfonated polyethylene, Vol. 8,
Ethylene–acrylic elastomers, Vol. 8,
Ethylene–propylene—diene rubber, Vol. 8,
Fluorocarbon elastomers, Vol. 8,
Nitrile rubber, Vol. 8,
Phosphazenes, Vol. 8,
Polybutadiene, Vol. 8,
Polychloroprene, Vol. 8,
Polyethers, Vol. 8,
Polyisoprene,
Thermoplastic elastomers,

POLYISOPRENE

The preparation of a synthetic polyisoprene was first reported in 1879 by Bouchardat (1), who treated isoprene [78-79-5] obtained from the destructive distillation of natural rubber with hydrochloric acid. This discovery led to a search for a way of converting isoprene into a material duplicating natural rubber (*Hevea brasilienses*). During World War II, scientists extensively studied the polymerization of isoprene with the hope of replicating natural rubber since the United States was temporarily cut off from sufficient natural rubber supplies. These studies were not successful. Finally, in 1954 the B.F. Goodrich Co. was successful in preparing a synthetic *cis*-1,4-polyisoprene [9003-31-0] through use of the newly discovered Ziegler transition-metal halide coordination-type catalyst (2,3). This catalyst consisted of a trialkylaluminum and titanium tetrachloride. Soon afterward, the Firestone Tire and Rubber Co. revealed a synthesis of *cis*-1,4-polyisoprene with a catalyst based on lithium metal (4). This catalyst yielded a polyisoprene with about 92% *cis*-1,4-structure. After achieving the goal of a synthetic *cis*-1,4-polyisoprene duplicating the structure and properties of natural rubber, workers focused on development of economical processes for commercial manufacture of isoprene monomer and for its commercial polymerization. In 1959–1960 the Shell Oil Co. came on stream with the first commercial plant for producing synthetic *cis*-1,4-polyisoprene using a lithium catalyst. Shortly thereafter, The Goodyear Tire & Rubber Co. began commercial production of *cis*-1,4-polyisoprene using a Ziegler catalyst. Later, in 1967, the B. F. Goodrich Co. began commercial production of synthetic polyisoprene using a Ziegler catalyst.

Physical Properties

In Table 1 some of the properties of raw synthetic *cis*-1,4-polyisoprene (Goodyear's Natsyn) and natural rubber (Hevea) are presented along with references that contain additional thermal, optical, electrical, and mechanical property data. Some properties of synthetic *trans*-1,4-polyisoprene (Kuraray TP-301) are also given. Molecular weights and mol wt distribution are determined by gel-permeation chromatography (gpc) (11).

Table 1. Properties of Polyisoprenes

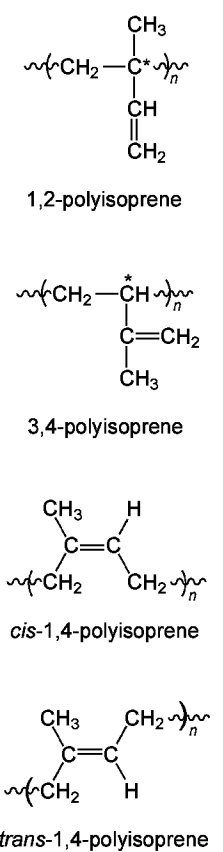
Property	<i>cis</i> -1,4-Polyisoprene ^a		<i>trans</i> -1,4-Polyisoprene ^b	
	Natural rubber (Hevea)	Natsyn 2200 ^c	Kuraray TP-301 ^d	Natural balata
density, g/mL	0.913	0.91	0.96	
refractive index, <i>n</i> _D	1.5191			
Mooney viscosity, ML 1 + 4, 100°C		70–90	30	25–33
ash content, % max	0.5–1.5	0.60	0.15	0.5
<i>T</i> _g , °C	–72	–72		
<i>T</i> _m , °C			67	67
<i>M</i> _w × 10 ^{–5}		7.55–9.55		
<i>M</i> _n × 10 ^{–5}		2.4–3.5		
microstructure, nmr				
<i>cis</i> -1,4, mol %	~100%	98.3 ^e		
<i>trans</i> -1,4, mol %		1.4	99	~100
3,4, mol %		0.3	0.2	0.2

^a Ref. 5, 6, 7, 8.
^b Ref. 9.
^c The Goodyear Tire & Rubber Co., Ziegler Ti–Al catalyst.
^d Gel content = 0%; volatile matter = 0.3%.

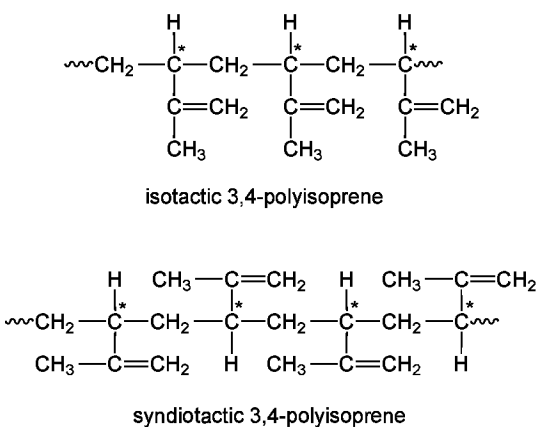
^e Cariflex, from Shell Nederland Chemie, and an alkylolithium catalyst (10), has 90.9% *cis*-1,4, 5.2% *trans*-1,4, and 3.9% 3,4.

Chemical Properties

Polymer Structure. Isoprene can undergo 1,4-, 1,2-, or 3,4-addition polymerization depending on the catalyst type and conditions, resulting in several structures:



For the 1,2- and 3,4-addition, a chiral carbon (marked by an asterisk) is formed which has an *R* or *S* configuration, but there is no net optical activity, because equal amounts of the *R* and *S* configurations are formed. The *R* and *S* configurations along the polymer chains lead to diastereomeric isomers called isotactic, syndiotactic, and atactic. In isotactic polyisoprene all monomer units have the same configuration as illustrated for isotactic 3,4-polyisoprene. In syndiotactic polyisoprene, the monomer units alternate between *R* and *S* configurations as illustrated for syndiotactic 3,4-polyisoprene.



In atactic polyisoprene, *R* and *S* configurations occur at random. The different modes of enchainment of isoprene (1,2-, 3,4-, *cis*-1,4-, and *trans*-1,4-) and the different stereochemical configurations that are possible result in eight possible stereoregular polyisoprenes. The best known polyisoprenes of high stereoregularity are *cis*-1,4-polyisoprene and *trans*-1,4-polyisoprene. Natural rubber (Hevea, Guayule, tree rubber) and synthetic *cis*-1,4-polyisoprene (Natsyn), currently produced in the United States by The Goodyear Tire & Rubber Co., are examples of *cis*-1,4-polyisoprene. Natural balata and gutta-percha and synthetic *trans*-1,4-polyisoprene, currently produced by the Kuraray Co., Ltd. of Japan, are examples of *trans*-1,4-polyisoprene. A 3,4-polyisoprene of 99–100% 3,4 units (12) has been synthesized. The *cis*-1,4-polyisoprene is a soft rubbery material, whereas the *trans*-1,4-polyisoprene is a hard, crystalline material.

The microstructures of several synthetic *cis*-1,4-polyisoprenes and natural rubber as determined by ¹³C-nmr (8,13) are shown in Table 1. ¹H-nmr measurements may also be used for microstructure characterization (8,13–15).

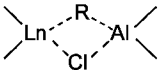
Polymerization

Ziegler-Natta. In 1954, the B. F. Goodrich Co. used the newly discovered Ziegler catalyst for polymerizing ethylene to polyethylene as a catalyst for making *cis*-1,4-polyisoprene (2,3). The catalyst consisted of titanium tetrachloride and a trialkylaluminum. Aliphatic or aromatic hydrocarbons were used as the polymerization solvents with the exclusion of oxygen and moisture. The method of preparing the Ziegler catalyst and polymerization variables, such as temperature and catalyst level, has been extensively studied (16,17). The active titanium species is the finely divided brown solid precipitate of β-TiCl₃ (16,18). The α-, δ-, and γ-crystalline forms of titanium trichloride do not give the *cis*-1,4-microstructure (19) but rather the *trans* polymer. In a hydrocarbon solvent the two catalyst components, triisobutylaluminum and titanium tetrachloride, react in the absence of monomer to yield a preformed catalyst which was found to be more reactive and gave more reproducible polymerizations than catalysts prepared *in situ* (20). A molar ratio of

Al:Ti in the range of 0.9–1.0 appeared optimum for *cis*-1,4-polyisoprene yield (20). Other factors such as catalyst preparation temperature, influence of the R group in the alkylaluminum compound (R₃Al), and catalyst aging have been extensively studied (16,17). Another variable studied was the effect of electron donors such as ethers and amines on the catalyst (21–24). The optimum molar ratio of ether to titanium was found to be in the range of 0.7 to 0.9, depending on the ether used (16,25). Modification of the catalyst with aromatic ethers resulted in high activity, a low isoprene oligomer content, and a low polymer gel content (25,26). The polymerization temperature and catalyst concentration directly influenced the polymer molecular weight and polymerization rate, but apparently did not affect the polymer microstructure. A reduction in the polymerization temperature from 50 to 0°C gave an approximate doubling of the polymer molecular weight as measured by dilute solution viscosity (dsv) (16). The monomer conversion to polymer was directly dependent on the catalyst level (27). However, the polymer molecular weight as measured by dsv was only approximately doubled by a twentyfold decrease in catalyst (26,28).

High (95–97%) *cis*-1,4-polyisoprene has been prepared with a catalyst consisting of an aluminum hydride derivative (alane) and titanium tetrachloride by the SNAM-Laboratori Riuniti Studi e Ricerche S. Donato Milanese in Milan, Italy. This catalyst does not contain a direct metal–carbon bond (29). With titanium tetrachloride and AlHCl₂·O(C₂H₅)₂, a halogen–alane complexed with diethyl ether, in benzene at 5°C, and an Al:Ti molar ratio of 1.6, a 96.5% *cis*-1,4-polyisoprene (by ir analysis) was obtained (29). Other alane-type catalysts contain trialkylamines, NR₃, as the Lewis bases: AlHCl₂·NR₃, AlH₂Cl·NR₃, AlH₃·NR₃ (30). Also, poly(*N*-alkyliminoalanes), (AlHNR)_{*n*}, in which R is alkyl or phenyl and *n* = 4–10, have been utilized with titanium tetrachloride (30). Cyclic polyiminoalanes (PIA), (AlHNR)_{*n*}, and titanium tetrachloride appeared to give the best performance in terms of activity and polymer molecular weight (31). The polyiminoalanes have been characterized as cyclic oligomers containing 4–10 iminic units.

Another catalyst system for preparing high *cis*-1,4-polyisoprene is based on rare-earth compounds in the lanthanide series, especially neodymium compounds. Chinese workers (32) reported the preparation of a polyisoprene containing 95% *cis*-1,4 units with Ln(naph)₃–(*i*-C₄H₉)₃Al–(C₂H₅)₃Al₂Cl₃ catalyst, where Ln is a rare-earth metal ion and naph is naphthenate. The highest activity catalyst was obtained with neodymium naphthenate which yielded polyisoprene in 96% conversion with 95% *cis*-1,4 and 5% 3,4 microstructure. Preparation of >94% *cis*-1,4-polyisoprene with lanthanide–trialkylaluminum catalysts such as Nd(naph)_{*n*}Cl_{3–*n*}, where *n* = 1 or 2; NdCl₃·3ROH; and neodymium trichloride phosphonate was also reported (33). The postulated active species of these catalysts was a bimetallic complex containing an alkylated rare-earth compound:



Polyisoprenes of 94–98% *cis*-1,4 content were obtained with lanthanum, cerium, praseodymium, neodymium, and other rare-earth metal ions (eg, LnCl₃) with trialkylaluminum (R₃Al) (34). Also, a NdCl₃·2THF(C₂H₅)₃Al catalyst has been used to prepare 95% *cis*-1,4-polyisoprene (35). *cis*-1,4-Polyisoprene of 98% *cis*-1,4 and 2% 3,4 content was obtained with organoaluminum–lanthanide catalysts, NdCl₃·*n*L, where L is an electron-donor ligand such as ethyl alcohol or butyl alcohol, or a long-chain alcohol, and *n* is 1 to 4 (36).

Amorphous (most likely atactic) 3,4-polyisoprene of 94–100% 3,4-microstructure was prepared with a (C₂H₅)₃Al–Ti(O–*n*-C₃H₇)₄ catalyst (11). Crystalline 3,4-polyisoprene containing about 70% 3,4-units and about 30% *cis*-1,4-microstructure was prepared using a catalyst derived from iron acetyl acetonate, trialkylaluminum, and an amine in benzene (37). However, this polyisoprene contained gel and was obtained in poor yield. Essentially gel-free crystallizable 3,4-polyisoprene of 70–85% 3,4-microstructure with the remainder being *cis*-1,4 microstructure was prepared in conversions of greater than 95% with a water-modified trialkylaluminum, ferric acetyl acetonate, and 1,10-phenanthroline catalyst (38). The 3,4-polyisoprene is stereoregular and believed to be syndiotactic or isotactic.

In 1955, *trans*-1,4-polyisoprene having a structure identical to that of natural gutta-percha was prepared using a triethylaluminum- α -titanium trichloride catalyst (39). Since then, vanadium catalysts, especially vanadium trichloride, with trialkylaluminums have been found to be the most active and stereospecific, and give the highest yields of *trans*-1,4-polyisoprene (17). Polymerization of isoprene in heptane using trialkylaluminium–vanadium trichloride catalyst yielded very high *trans*-1,4-polyisoprene (40), and a triethylaluminum–vanadium trichloride catalyst gave about 97% *trans*-1,4 microstructure (nmr analysis) (41). Modification of the trialkylaluminum–vanadium trichloride catalyst with ethers, such as diisopropyl ether, increased the polymer yield (42). Increased polymer yields have also been obtained with vanadium trichloride deposited on inert supports, eg, kaolin (43).

Alfin Catalysts. Alfin catalysts (44,45) give polyisoprenes of high *trans*-1,4 microstructure (46). For example, a typical Alfin catalyst gives polyisoprene of 52% *trans*-1,4, 27% *cis*-1,4, 16% 3,4, and 5% 1,2 content (ir analysis) (46). One type of Alfin catalyst consists of allylsodium, sodium isopropoxide, and sodium chloride (47,48). Because of the mixed microstructure polyisoprene produced, Alfin catalysts are not used commercially.

Alkali Metal Catalysts. The polymerization of isoprene with sodium metal was reported in 1911 (49,50). In hydrocarbon solvent or bulk, the polymerization of isoprene with alkali metals occurs heterogeneously, whereas in highly polar solvents the polymerization is homogeneous (51–53). Of the alkali metals, only lithium in bulk or hydrocarbon solvent gives over 90% *cis*-1,4 microstructure. Sodium or potassium metals in *n*-heptane give no *cis*-1,4 microstructure, and 48–58 mol % *trans*-1,4, 35–42% 3,4, and 7–10% 1,2 microstructure (46). Alkali metals in benzene or tetrahydrofuran with crown ethers form solutions that readily polymerize isoprene; however, the 1,4 content of the polyisoprene is low (54). For example, the polyisoprene formed with sodium metal and dicyclohexyl-18-crown-6 (crown ether) in benzene at 10°C contains 32% 1,4-, 44% 3,4-, and 24% 1,2-isoprene units (54).

High *cis*-1,4-polyisoprene was prepared using finely divided lithium metal in petroleum jelly as catalyst in 1956 (55). The bulk polymerization started after an induction period, and the polymerization mass thickened and became solid. Polyisoprene of low gel, high inherent viscosity, and up to 94% *cis*-1,4 content (the remainder was 3,4) was obtained. A lithium metal dispersion in hexane or heptane also gave high *cis*-1,4-polyisoprene (46,56), whereas lithium metal dispersions in polar solvents such as tetrahydrofuran (THF), diethyl ether, and dioxane gave essentially no *cis*-1,4 microstructure (46). For example, in THF polyisoprene of 34 mol % *trans*-1,4-, 52% 3,4-, and 14% 1,2- was produced (46). In the late 1950s, the lithium dispersions were generally replaced by organolithium compounds, which were easier to handle, gave homogeneous systems, and could be used without obtaining the poorly reproducible induction times that attended use of lithium metal.

Anionic Polymerization. Polyisoprenes of predictable molecular weight and uniform chain length can be prepared by anionic polymerization, which is characterized by living growing chains and the absence of chain-termination (57–61). Therefore, polyisoprenes of predictable molecular weight can be prepared from simple stoichiometry. Very narrow molecular weight distributions, approaching the Poisson distribution, are possible. Organolithiums, especially the butyllithiums, are most frequently utilized as initiators of the anionic polymerization of isoprene since they are soluble in both polar and nonpolar solvents. Both *sec*- and *tert*-butyllithiums have higher initiation rates than that of *n*-butyllithium (16). The polymerization solvent profoundly influences the polyisoprene microstructure as shown in Table 2 (62,63).

Table 2. Effect of Polymerization Solvent on Polyisoprene Microstructure

Initiator	Solvent	Temperature, °C	Mol % ^a		
			<i>cis</i> -1,4	<i>trans</i> -1,4	3,4
<i>n</i> -C ₄ H ₉ Li	cyclohexane	30	80	15	5
<i>sec</i> -C ₄ H ₉ Li	benzene	20	71	23	6
<i>sec</i> -C ₄ H ₉ Li	THF	30	0	69	31

^a Precision of 300 MHz ¹H nmr: *cis*-1,4 or *trans* – 1,4 = ±2%; 3,4 or 1,2 = ±1%. 1,2 measured 0% in all three cases.

A change from an aliphatic or aromatic hydrocarbon solvent (cyclohexane, benzene) to a polar solvent (THF) leads to a large increase in *trans*-1,4 and 3,4 microstructure (58). Organolithium compounds are highly associated; *sec*-butyllithium in benzene or cyclohexane exists as a tetramer, and *n*-butyllithium as a hexamer (64,65). This association in hydrocarbon solvents results partly in the slow initiation observed between some organolithiums and isoprene (66). At low initiator concentrations, the polymerization rate of isoprene in alkyllithium polymerization is proportional to monomer and alkyllithium concentrations (67). 3,4-Polyisoprenes are obtained by modification of the lithium polymerization with ethers, such as the dialkyl ethers of ethylene glycol or tertiary amines (68,69).

Cationic Polymerization. A small amount of isoprene is cationically copolymerized with isobutylene in the commercial process for making butyl rubber, wherein the isoprene provides the unsaturation required for sulfur vulcanization. Homopolymerization of isoprene by cationic catalysts can lead to cyclized products and loss of unsaturation (70,71), as for example, during polymerizations initiated by boron trifluoride (72). Cationic polymerization of isoprene with BF₃, SnCl₄, or AlCl₃ catalysts in pentane, chloroform, or ethylbenzene from −78 to 30°C at around 50% conversion gives about 90% *trans*-1,4-polyisoprene structure; the balance is 1,2 and 3,4 microstructure (no *cis*-1,4 microstructure) (73). An insoluble powder was formed by cyclopolymerization of isoprene using titanium tetrachloride/Grignard reagent (74) or ethylaluminum dichloride (75).

Free-Radical Polymerization. The best method for polymerizing isoprene by a free-radical process is emulsion polymerization. Using potassium persulfate [7727-21-1] as initiator at 50°C, a 75% conversion to polyisoprene in 15 h was obtained (76). A typical emulsion polymerization recipe is given as follows (77).

Ingredient	Parts by wt
water	180
isoprene	100
potassium fatty acid soap	5
potassium chloride	1
potassium persulfate	0.15–0.60
<i>tert</i> -dodecyl mercaptan	0.1–0.8

The predominant microstructure obtained by emulsion polymerization is *trans*-1,4, determined by ir, as shown in mol % for two temperatures and initiators (72).

Initiator	<i>cis</i> -1,4	<i>trans</i> -1,4	3,4	1,2
cumene hydroperoxide at 5°C	1.7	86.2	5.4	6.7
potassium persulfate at 50°C	17.6	71.9	5.3	5.2

Since emulsion polyisoprene has low *cis*-1,4 microstructure, it has poorer physical properties than the high *cis*-1,4-polyisoprene and has not been commercialized.

Manufacture and Processing

Currently, *cis*-1,4-polyisoprene is manufactured in the United States only by The Goodyear Tire & Rubber Co. at Beaumont, Texas. In Western Europe, only the Shell Nederland Chemie (Royal Dutch/Shell Group) commercially produces *cis*-1,4-polyisoprene; in Japan, the producers are Japan Synthetic Rubber Co., Ltd. and Nippon Zeon Co., Ltd. The sole world producer of *trans*-1,4-polyisoprene is Kuraray Co., Ltd. in Japan.

Dry prepurified isoprene monomer and dry aliphatic hydrocarbon solvent along with a Ziegler catalyst consisting of triisobutylaluminum and titanium tetrachloride are utilized in one commercial solution process for *cis*-1,4-polyisoprene. Isoprene, the Ziegler catalyst, and recycled aliphatic hydrocarbon solvent in an inert atmosphere (nitrogen) are continuously fed into a continuous polymerization reactor system, and the isoprene is polymerized to the desired conversion. As the polymerization stream leaves the reactor system, a shortstop and an antioxidant are added to stop the polymerization and protect the polyisoprene throughout the finishing operation. The shortstopped polymerization stream then proceeds to the steam stripping operation, where the unreacted isoprene monomer and aliphatic hydrocarbon solvent are removed. After separation and purification, the isoprene and aliphatic hydrocarbon solvent are recycled back to the beginning of the process. The rubber crumb from the stripping operation is fed through dewaterers and dryers to obtain dry rubber crumb. The dried rubber crumb is baled and packaged in 50 μm (2.0 mil) polyethylene film. The bales weigh about 34 kg. Modern polymerization plants such as Goodyear's at Beaumont, Texas employ the latest process design and computer controls for efficient operation. An on-site cogeneration power plant allows efficient use of high pressure steam to run equipment and provide a large portion of the plant's electrical needs. Air emissions and wastewater are treated to meet environmental laws and regulations. Over 95% of the liquid wastes generated by the plant are reused (78). Detailed flow sheets of commercial plants have been published (79,80). The microstructures of two commercial *cis*-1,4-polyisoprenes (Natsyn and Cariflex) prepared with two different catalysts and natural rubber are compared in Table 1.

Compounding. Another important part of the use of polyisoprene rubber is its compounding into formulations that can be utilized to make finished goods and articles. The raw polyisoprene rubber is mixed in a Banbury (internal) mixer or on a mill with ingredients such as fillers (carbon black, clays, or silicates), plasticizers, antioxidants, accelerators, and sulfur. The filler, eg, carbon black, gives durability and reduces cost of the compound. The sulfur functions as a cross-linking agent for the highly unsaturated polyisoprene. The accelerators, eg, sulfenamides, thiazoles, guanidines, thiuram sulfides, and thiocarbamates, help reduce the vulcanization time. The antioxidants (qv), eg, arylamines, hindered phenols, etc, retard the deterioration of the finished article due to aging. Antiozonants (qv), eg, *p*-phenylenediamine, and waxes are also used in tire formulations to retard cracking under stress. The resulting rubber compound is calendered, molded, extruded, or fabricated into the desired shape and vulcanized or cured at around 150°C.

In Tables 3 and 4 typical formulations and properties are shown for black-filled stocks of natural rubber and synthetic *cis*-1,4-polyisoprene and a blend of the two. Although black-filled stocks of synthetic polyisoprene and natural rubber exhibit many similar properties, the tackiness and green strength of synthetic polyisoprene are inferior. However, synthetic polyisoprene exhibits a uniform cure rate that can lead to automation in production of rubber parts. Blends of synthetic polyisoprene and natural rubber can smooth out processing or cure variations of pure natural rubber. Synthetic polyisoprene is also blended with other general-purpose elastomers such as *cis*-1,4-polybutadiene and styrene–butadiene rubber (SBR) to obtain combinations of properties not available with a single elastomer. For example, the tear and tensile strengths of a polyisoprene compound can be increased by blending with *cis*-polybutadiene and SBR. Compounding information and formulations are available from polymer and rubber chemical producers (5,6,81).

Table 3. Black-Loaded Formulation for Natsyn and Natural Rubber^a

Ingredient	Compound 1, phr	Compound 2, phr	Compound 3, phr
Natsyn 2200	100.00		50.00
natural rubber SMR-L		100.00	50.00
zinc oxide	3.00	3.00	3.00

stearic acid	2.00	1.00	1.50
Wingstay 100 ^b	1.25	0.25	0.75
N-330 carbon black	40.00	40.00	40.00
naphthenic oil	5.00	5.00	5.00
Santocure ^c	1.20	1.20	1.20
Unads ^d	0.20	0.10	0.15
sulfur	2.00	2.00	2.00
<i>Total</i>	<i>154.65</i>	<i>152.55</i>	<i>153.60</i>

^a Ref. 5. Best cure, 150°C for 20 min; phr by weight.

^b Mixed diaryl-*p*-phenylenediamines, antioxidant of The Goodyear Tire & Rubber Co.

^c Accelerator of Monsanto Co.

^d Tetramethylthiuram monosulfide, accelerator of R. T. Vanderbilt Co.

Table 4. Properties of Black-Filled Stocks for Natsyn and Natural Rubber^a

Property	Compound 1	Compound 2	Compound 3
tensile strength, MPa ^{b,c}	23.1	25.9	26.2
elongation, %	460	480	490
300% modulus, MPa ^c	12.6	12.4	12.8
Shore A hardness	68	65	66
tear strength, kN/m ^d	58.1	72.3	60.6
compression set, 22 h at 70°C, %	10.9	16.6	10.9
Goodyear-Healey rebound pendulum test ^e			
cold rebound, %	74.6	79.8	78.6
hot rebound, %	85.9	87.7	87.1
MST abrasion resistance	119	168	165
green strength			
stress at yield, MPa ^c	0.28	0.38	0.33
ultimate tensile, MPa ^c	0.43	2.35	1.41
ultimate elongation, %	1950	630	760

^a Ref. 5. Formulation of Table 3.

^b Scott tensile, die C, 50 cm/min.

^c To convert MPa to psi, multiply by 145.

^d To convert kN/m to ppi, multiply by 5.71.

^e ASTM D1054, Ref. 82.

Economic Aspects

In 1993, synthetic *cis*-1,4-polyisoprene was manufactured commercially in the United States only by The Goodyear Tire and Rubber Co. at Beaumont, Texas. This plant has a capacity of 61,000 metric tons/yr and utilizes a Ziegler catalyst (83). Since 1985, annual production of polyisoprene has been flat at about 48,000 to 49,000 metric tons (84). In 1990, Bayer AG purchased Polysar, Inc. and the former B. F. Goodrich Co. polyisoprene plant at Orange, Texas which is on inactive status and has an estimated potential annual capacity of 55,000 metric tons. The only Western European producer of synthetic *cis*-1,4-polyisoprene is Shell Nederland Chemie (member of the Royal Dutch/Shell Group) with a plant at Pernis, (near Rotterdam) Netherlands with an annual estimated plant capacity of 75,000 metric tons in 1983 (84,85). The plant capacity has since been reduced to about 45,000 metric tons per year (83). In 1989, the estimated production of *cis*-1,4-polyisoprene in Western Europe was 20,000 metric tons (84). In 1989 in Japan, the annual production capacity for synthetic *cis*-1,4-polyisoprene was about 72,000 metric tons with Japan Synthetic Rubber Co., Ltd. having an annual capacity of 35,000 metric tons, and Nippon Zeon Co., Ltd. having an annual capacity of 37,000 metric tons. In 1989, the estimated production of *cis*-polyisoprene in Japan was 55,000 metric tons (84). In 1992, Coperbo in Brazil, South America was to start up a plant with an estimated annual capacity of 35,000 metric tons/yr (83,84). In 1990, in the USSR and other European Centrally Planned Economy Countries, the total plant capacity for synthetic polyisoprene was estimated to be about six times as large as the rest of the world. In the USSR, the production capacity was about 960,000 metric tons/yr, and in 1989, an estimated 770,000 metric tons were produced (84). In mid-1992, the former Soviet Union (Commonwealth of Independent States) was expected to reduce its output and capacity for synthetic polyisoprene to about 75% of its current level (86).

trans-1,4-Polyisoprene is not produced commercially in the United States, although the 1990 consumption was estimated to be around 2000 metric tons (synthetic and natural) (84). The *trans*-polyisoprene is used mainly for golf ball covers for premium golf balls and for medical applications. In Japan, Kuraray Co., Ltd. produces about 400 metric tons/yr of *trans*-1,4-polyisoprene (84).

In 1989, the estimated United States consumption of synthetic polyisoprene was 51,000 metric tons, and of natural rubber 890,000 metric tons (84). The price for synthetic *cis*-1,4-polyisoprene, namely, Natsyn 2200, Natsyn 2205, or Natsyn 2210 in the United States in August, 1992, was \$1.98/kg in carload or truckload quantities fob at Beaumont, Texas (87). The price of synthetic *trans*-1,4-polyisoprene from Kuraray Co., Ltd. in the United States was \$21.80/kg in 545-kg lots (88).

Specifications, Standards, Quality Control

Specifications for solid *cis*-1,4-polyisoprenes are shown in Table 5 and include analyses for volatile matter, extractables, ash, and Mooney viscosity at 100°C. Standard method ASTM D1416 includes chemical analysis methods for volatile matter, extractables, and total ash, while ASTM D1646 includes Mooney viscosity (82). The Monsanto rheometer data for vulcanizates prepared by a standard recipe may also be specified. This formulation for vulcanizate (ASTM D3403) is mixed in a Banbury mixer in two passes with all but sulfur and accelerator added in first pass:

Material	Phr
Natsyn polyisoprene	100.0
HAF black	35.00
stearic acid	2.00

zinc ozide	5.00
Santocure NS	0.70
sulfur	2.25
<i>Total</i>	<i>144.95</i>

Table 5. Specifications of *cis*-1,4-Polyisoprenes^a

Property	Natsyn 2200	Natsyn 2205 ^b	Natsyn 2210	ASTM test method
raw polymer				
volatile matter, % max	0.5	0.5	0.5	D1416
extractables, % max	3.0	3.0	3.0	D1416
ash, % max	0.6	0.6	0.6	D1416
Mooney, ML-4 at 100°C	70–90	70–90	50–65	D1646
vulcanizate ^c torque ^d N·m ^e				
minimum	0.61–0.90	0.61–0.90	0.51–0.85	
maximum	3.78–4.30	3.78–4.30	3.38–4.17	
<i>t</i> _f , min ^f	5.8–8.2	5.8–8.2	5.8–8.8	
<i>t</i> ₅₀ , min	8.7–12.3	8.7–12.3	8.2–12.7	
<i>t</i> ₉₀ , min ^g	12.8–16.8	12.8–16.8	12.4–16.9	

^a Goodyear Chemicals Product Specification Sheets: Natsyn 2200 (Sept. 17, 1991); Natsyn 2205 (Sept. 17, 1991); Natsyn 2210 (May 9, 1990).

^b Gel particles >1.0 mm, none; 0.5 to 1.0 mm, 4 max.

^c Formulation for vulcanizate ASTM D3403.

^d Monsanto rheometer at 150°C, 1 deg arc, 1.7 Hz, 30 min chart, ASTM D2084.

^e To convert Nm to in.-lb., multiply by 8.85.

^f Rise above minimum torque with 1 arc.

^g Time to maximum torque.

Health and Safety Factors

Polyisoprene rubber is relatively nonhazardous, but must be kept away from sparks, open flames, or excessive heat because it will burn. The current Material Safety Data Sheet (MSDS) should always be checked for known hazards before using polyisoprene or any other chemical materials.

Uses

cis-1,4-Polyisoprene is used in tires and tire products, belts, gaskets, hoses, foam rubber, molded and mechanical goods, bottle nipples, gloves, caulking, sealants, footwear and sporting goods, rubber bands, erasers, and rubber sheeting. Synthetic *cis*-1,4-polyisoprene and natural rubber have the same type of applications because of their similar structure and properties. About 60% of the synthetic *cis*-1,4-polyisoprene was used in tires and tire products in 1989 (84,89), and about 12% in mechanical goods. Footwear, especially sports footwear, accounted for about 4% of the polyisoprene consumption in 1989; all other applications accounted for 24%. Both synthetic *cis*-1,4-polyisoprene and natural rubber are used in construction of radial passenger car tire carcasses because of their good green strength and building tack. However, polyisoprene is not normally used in passenger car tire treads. Generally, passenger car tire treads consist of blends of polybutadiene and styrene–butadiene rubber (SBR) (84,89). The mechanical goods application area includes molded automotive goods, dairy industry products, and medical applications. Also, synthetic *cis*-1,4-polyisoprene is used in heavy-duty truck and bus tire treads where outstanding treadwear and low heat buildup under heavy loads are important.

Since *trans*-1,4-polyisoprene is a crystalline thermoplastic, it resists abrasion, scuffing, and cutting, and is used mainly in high quality golf ball covers and orthopedic devices and splints, and to some extent in transmission belts, cable coverings, and adhesives (84). *trans*-1,4-Polyisoprene is used in prosthetics, splints, braces, casts, and attachments for artificial limbs because it possesses an unusual combination of properties (90). *trans*-Polyisoprene is a tough, rigid, durable, and lightweight material at room temperature, but when immersed in hot water, it softens and does not crystallize immediately as it cools to body temperature, thereby permitting molding directly to a patient's body.

3,4-Polyisoprene is finding some applications in specialty tire rubber compounds (69,91).

BIBLIOGRAPHY

"Elastomers, Synthetic (Polyisoprene)" in *ECT* 3rd ed., Vol. 8, pp. 582–592, by H. Tucker and S. E. Horne, Jr., B.F. Goodrich Co.

1. G. Bouchardat, *Compt. Rend.* **89**, 1117 (1879).
2. **Belg. Pat. 543,292** (June 2, 1956), (to Goodrich-Gulf Chemical Co.).
3. S. E. Horne and co-workers, *Ind. Eng. Chem.* **48**, 784 (1956).
4. F. W. Stavely and co-workers, *Ind. Eng. Chem.* **48**, 778 (1956).
5. *Natsyn Polyisoprene Rubber Bulletin*, Goodyear Chemical Division, The Goodyear Tire & Rubber Co., Akron, Ohio, Oct., 1988.
6. K. A. Valin in R. F. Ohm, ed., *The Vanderbilt Rubber Handbook*, R. T. Vanderbilt Co., Inc., Norwalk, Conn., 1990, p. 44.
7. L. A. Wood in J. Brandrup and E. H. Immergut, eds., *Polymer Handbook*, 3rd ed., John Wiley & Sons, Inc., New York, 1989, pp. V/7 and V/8.
8. D. R. Burfield and Y. Tanaka, *Polymer* **28**, 907 (1987).
9. *Kuraray TP Technical Information No. 731, Kuraray Trans Polyisoprene TP-301*, Kuraray Isoprene Chemical Co., Ltd., Synthetic Rubber Dept., 3-8-2, Nihonbashi, Chuo-Ku, Tokyo, Japan, Oct. 1, 1982.
10. A. F. Halasa, *Rubber Chem. Technol.* **55**, 253 (1982).
11. J. Cazes, *J. Chem. Ed.* **43**, A567 (1966).
12. G. Natta, L. Porti, and A. Carbonaro, *Makromol. Chem.* **77**, 126 (1964).
13. Y. Tanaka, *Rubber Chem. Technol.* **64**, 325 (1991).
14. H. Sato, A. Ono, and Y. Tanaka, *Polymer* **18**, 580 (1977).
15. H. Sato and Y. Tanaka, *J. Polym. Sci., Polym. Chem. Ed.* **17**, 3551 (1979).
16. E. Schoenberg and co-workers, *Rubber Chem. Technol.* **52**, 526 (1979).
17. L. Porri and A. Giarrusso in G. Allen and J. C. Bevington, eds., *Comprehensive Polymer Science*, Vol. 4, Pergamon Press, Inc., Elmsford, N.Y., 1989,

- pp. 53–108.
18. G. Natta and co-workers, *Rend. Acad. Naz. Lincei* **24**(8), 121 (1958).
 19. G. Natta, L. Porri, and L. Fiore, *Gazz. Chim. It.* **89**, 761 (1959).
 20. E. Schoenberg, D. L. Chalfant, and R. H. Mayor, *Rubber Chem. Technol.* **37**, 103 (1964).
 21. T. L. Brown, *Adv. Organomet. Chem.* **3**, 365 (1965).
 22. **U.S. Pat. 3,165,503** (Jan. 12, 1965), H. Kahn and S. E. Horne, Jr. (to Goodrich-Gulf, Inc.).
 23. **Ger. Pat. 1,301,487** (Nov. 20, 1969), (to Chemische Werke Huls, AG).
 24. **Belg. Pat. 726,446** (June 16, 1969), (to Farbenfab, Bayer AG).
 25. E. Schoenberg and co-workers, *Metal Alkyl Catalyzed Polymerizations*, ACS Symposium Series, Div. Ind. Eng. Chem, New York, Sept., 1963.
 26. W. M. Saltman, F. S. Farson, and E. Schoenberg, *Rubber and Plastics Age (London)* **46**, 502 (1965).
 27. C. F. Gibbs and co-workers, *Rubber World* **144**(1), 69 (1961).
 28. S. Y. Bresler and co-workers, *Rubber Chem. Technol.* **33**, 689 (1960).
 29. W. Marconi and co-workers, *Makromol. Chem.* **71**, 118 and 134 (1964).
 30. A. Mazzei, S. Cucinella, and W. Marconi, *Makromol. Chem.* **122**, 168 (1969).
 31. A. Balducci and co-workers, *Rubber Chem. Technol.* **48**, 736 (1975).
 32. Academia Sinica, *Sci. Sin.* **17**, 656 (1974).
 33. J. Ouyang, F. Wang, and Z. Shen, *Proceedings of China-U.S. Bilateral Symposium on Polymer Chemistry and Physics*, Science Press, Van Nostrand Reinhold Co., New York, 1981, p. 382.
 34. Z. Shen and co-workers, *J. Polym. Sci., Polym. Chem. Ed.* **18**, 3345 (1980).
 35. J. H. Yang and co-workers, *Macromolecules* **15**, 230 (1982).
 36. H. L. Hsieh and H. C. Yeh, *Rubber Chem. Technol.* **58**, 117 (1985).
 37. Sun and Wang, *Acta Polymerica Sinica* **2**, 145 (1988).
 38. **U.S. Pat. 5,082,906** (Jan. 21, 1992), A. F. Halasa and W. L. Hsu (to The Goodyear Tire & Rubber Co.).
 39. **Ital. Pat. 536,631** (Dec. 7, 1955), G. Natta, L. Porri, and G. Mazzanti (to Montecatini SpA).
 40. G. Natta and co-workers, *Chim. Ind. Milan* **40**, 362 (1958); J. Boor, Jr., *Ziegler-Natta Catalysts and Polymerizations*, Academic Press, Inc., New York, 1979, pp. 145, 146, and 155.
 41. E. G. Lovering and W. B. Wright, *J. Polym. Sci., Part A-1* **6**, 2221 (1968).
 42. W. Cooper and co-workers, *J. Polym. Sci., Part C* **4**, 211 (1964).
 43. J. S. Lasky, H. K. Garner, and R. H. Ewart, *Ind. Eng. Chem. Prod. Res. Dev.* **1**, 82 (1962).
 44. A. A. Morton, E. E. Magat, and R. L. Letsinger, *J. Am. Chem. Soc.* **69**, 950 (1947).
 45. A. A. Morton, *Ind. Eng. Chem.* **42**, 1488 (1950).
 46. A. V. Tobolsky and C. E. Rogers, *J. Polym. Sci.* **40**, 73 (1959).
 47. V. L. Hansley and H. Greenberg, *Rubber Age* **94**, 87 (1963).
 48. A. A. Morton in N. M. Bikales, ed., *Encyclopedia of Polymer Science and Technology*, Vol. 1, 1st ed., Wiley-Interscience, New York, 1964, p. 629.
 49. **Brit. Pat. 24,790** (Oct. 25, 1911), F. E. Matthews and E. H. Strange.
 50. C. H. Harries, *Ann. Chem.* 383, 184 (1911).
 51. A. Essel and co-workers, *J. Polym. Sci. Polym. Chem. Ed.* **13**, 1847 (1975).
 52. A. Essel, R. Salle, and Q. T. Pham, *J. Polym. Sci. Polym. Chem. Ed.* **13**, 1869 (1975).
 53. R. Salle and co-workers, *J. Polym. Sci. Polym. Chem. Ed.* **13**, 1855 (1975).
 54. S. Alev. F. Schue, and B. Kaempf, *J. Polym. Sci. Polym. Lett. Ed.* **13**, 397 (1975).
 55. F. W. Stavely and co-workers, *Ind. Eng. Chem.* **48**, 778 (1956); *Rubber Chem. Technol.* **29**, 673 (1956).
 56. H. F. Diem, H. Tucker, and C. F. Gibbs, *Rubber Chem. Technol.* **34**, 191 (1961).
 57. J. E. McGrath, ed., *Anionic Polymerization, Kinetics, Mechanisms, and Synthesis*, ACS Symposium Series No. 166, American Chemical Society, Washington, D.C., 1981.
 58. M. Morton, *Anionic Polymerization: Principles and Practice*, Academic Press, Inc., New York, 1983.
 59. R. N. Young, R. P. Quirk, and L. J. Fetters in *Advances in Polymer Science*, Vol. 56, Springer-Verlag, New York, 1984, p. 1.
 60. M. Szwarc in *Advances in Polymer Science*, Vol. 49, Springer-Verlag, New York, 1983, p. 1.
 61. A. Noshay and J. E. McGrath, *Block Copolymers Overview and Critical Survey*, Academic Press, Inc., New York, 1977.
 62. D. J. Worsfold and S. Bywater, *Can. J. Chem.* **42**, 2884 (1964).
 63. F. Schue, D. J. Worsfold, and S. Bywater, *Macromolecules* **3**, 509 (1970).
 64. S. Bywater and D. J. Worsfold, *J. Organomet. Chem.* **10**, 1 (1967).
 65. M. Morton and R. A. Pett, and L. J. Fetters, *Macromolecules* **3**, 333 (1970).
 66. M. Morton and L. J. Fetters, *Rubber Chem. Technol.* **48**, 359 (1975).
 67. H. L. Hsieh, *J. Polym. Sci., Part A* **3**, 153 (1965).
 68. K. H. Nordsiek, *Kautsch. Gummi Kunstst.* **35**, 1032 (1982).
 69. K. H. Nordsiek and J. Wolpers, *Kautsch. Gummi Kunstst.* **43**(9), 755 (1990); **U.S. Pat. 4,894,425** (Jan. 16, 1990), W. Helleman, K. H. Nordsiek, and J. Wolpers (to Huels Aktiengesellschaft).
 70. W. Cooper in P. H. Plesch, ed., *The Chemistry of Cationic Polymerization*, The Macmillan Publishing Co., New York, 1963, p. 349.
 71. J. P. Kennedy, *Cationic Polymerization of Olefins: A Critical Inventory*, John Wiley & Sons, Inc., New York, 1975, pp. 164–166.
 72. W. S. Richardson and A. Sacher, *J. Polym. Sci.* **10**, 353 (1953).
 73. W. S. Richardson, *J. Polym. Sci.* **13**, 325 (1954).
 74. N. G. Gaylord and co-workers, *J. Am. Chem. Soc.* **85**, 641 (1963).
 75. I. Kossler, M. Stolka, and K. Mach, *J. Polym. Sci. Part C* **4**, 977 (1964).
 76. C. S. Marvel and J. L. R. Williams, *J. Polym. Sci.* **4**, 265 (1949).
 77. M. Morton, J. A. Cala, and I. Piirma, *J. Polym. Sci.* **15**, 167 (1955).
 78. *Goodyear-Beaumont Chemical Plant Booklet*, Goodyear Chemical Division, The Goodyear Tire & Rubber Co., Akron, Ohio, 1992.
 79. R. L. Magovern and S. L. Soder, *Polyisoprene, Report No. 82, Process Economics Program*, Stanford Research Institute, Menlo Park, Calif., Dec., 1972.
 80. *Hydrocarbon Proc.* **50**, 199–200 (Nov., 1971).
 81. M. Morton, *Rubber Technology*, 3rd ed., Van Nostrand Reinhold, New York, 1987, pp. 235–259; M. Morton, *Rubber Technology*, 2nd ed., Robert E. Krieger Publishing Co., Malabar, FL, 1973, pp. 274–301 and 19–113.
 82. R. A. Storer, ed., *1990 Annual Books of ASTM Standards*, Vol. 09.01, ASTM, Philadelphia, Pa., 1990, pp. 227, 307, Section 9.
 83. *Rubber Statistical Bulletin*, Vol. 45, No. 11, The International Rubber Study Group, Wembley, UK, Aug., 1991, p. 47.
 84. J. F. Auchter, *Polyisoprene Elastomers, Chemical Economics Handbook*, CEH Product Review, SRI International, Menlo Park, Calif., Jan., 1991.
 85. *Rubber Statistical Bulletin*, Vol. 37, No. 4, The International Rubber Study Group, London, Jan., 1983, p. 49.
 86. B. Davis, *Rubber & Plastics News* **21**(27), 6 (July 6, 1992).
 87. Technical data, The Goodyear Tire & Rubber Co., Akron, Ohio, Aug. 1, 1992.
 88. E. L. Puskas, Co., Akron, Ohio, Sales Agents and Distributors for Kuraray Co., Ltd., (Japan), *Trans-Polyisoprene TP-301*, Dec., 1991.
 89. B. F. Greek, *Chem. Eng. News* **69**(19), 37 (May 13, 1991).

90. E. J. Buckler, *Can. J. Chem. Eng.* **62**(1), 3 (1984).
91. H. Takino and co-workers, *Kautsch. Gummi Kunstst.* **43**(9), 761 (1990).

Michael Senyek
The Goodyear Tire & Rubber Co.



ELASTOMERS, SYNTHETIC-STYRENE–BUTADIENE RUBBER.

See STYRENE BUTADIENE RUBBER.

THERMOPLASTIC ELASTOMERS

Thermoplastic elastomers (TPE) were introduced in the 1960s. They have shown rapid growth since then and have been the subject of many conferences, symposia, etc. In particular, the developments in this field have been covered fairly recently in two books: one dealing primarily with the scientific aspects of these polymers (1) and the other dealing primarily with their applications (2). Thermoplastic elastomers have many of the physical properties of rubbers, ie, softness, flexibility, and resilience; but in contrast to conventional rubbers, they are processed as thermoplastics. Rubbers must be cross-linked to give useful properties. In the terminology of the plastics industry, vulcanization is a thermosetting process. Like other thermosetting processes, it is slow and irreversible and takes place upon heating. With thermoplastic elastomers, on the other hand, the transition from a processible melt to a solid, rubberlike object is rapid and reversible and takes place upon cooling. Thermoplastic elastomers can be processed using conventional plastics techniques, such as injection molding and extrusion; scrap is usually recycled.

Because of increased production and the lower cost of raw material, thermoplastic elastomeric materials are a significant and growing part of the total polymers market. World consumption in 1995 is estimated to approach 1,000,000 metric tons (3). However, because the melt to solid transition is reversible, some properties of thermoplastic elastomers, eg, compression set, solvent resistance, and resistance to deformation at high temperatures, are usually not as good as those of the conventional vulcanized rubbers. Applications of thermoplastic elastomers are, therefore, in areas where these properties are less important, eg, footwear, wire insulation, adhesives, polymer blending, and not in areas such as automobile tires.

The classification given in Table 1 is based on the process, ie, thermosetting or thermoplastic, by which polymers in general are formed into useful articles and on the mechanical properties, ie, rigid, flexible, or rubbery, of the final product. All commercial polymers used for molding, extrusion, etc, fit into one of these six classifications; the thermoplastic elastomers are the newest.

Table 1. Classification of Polymers According to Properties and Processing

Property	Thermoset	Thermoplastic
rigid	epoxy, melamine–formaldehyde, sheet molding compounds	polypropylene, high density polyethylene, polystyrene
flexible	highly vulcanized rubber	low density polyethylene, ethylene–vinyl, acetate copolymer, plasticized PVC
rubbery	vulcanized rubber	thermoplastic rubbers

Structure

Thermoplastic elastomers are often multiphase compositions in which the phases are intimately dispersed. In many cases, the phases are chemically bonded by block or graft copolymerization. In others, a fine dispersion is apparently sufficient. In these multiphase systems, at least one phase consists of a material that is hard at room temperature but becomes fluid upon heating. Another phase consists of a softer material that is rubberlike at RT. A simple structure is an A–B–A block copolymer, where A is a hard phase and B an elastomer, eg, poly(styrene-*b*-elastomer-*b*-styrene).

Most polymer pairs are thermodynamically incompatible and mixtures separate into two phases. This is true even when the polymeric species are part of the same molecule, as in these block copolymers. Poly(styrene-*b*-elastomer-*b*-styrene) copolymers, in which the elastomer is the main constituent, should have a structure similar to that shown in Figure 1. Here, the polystyrene end segments form separate spherical regions, ie, domains, dispersed in a continuous elastomer phase. Most of the polymer molecules have end polystyrene segments in different domains. At RT, these polystyrene domains are hard and act as physical crosslinks, tying the elastomer chains together in a three-dimensional network. In some ways, this is similar to the network formed by vulcanizing conventional rubbers using sulfur cross-links. The main difference is that in thermoplastic elastomers the domains lose their strength when the material is heated or dissolved in solvents. This allows the polymer or its solution to flow. When the material is cooled down or the solvent is evaporated, the domains harden and the network regains its original integrity. This explanation of the properties of thermoplastic elastomers has been given in terms of a poly(styrene-*b*-elastomer-*b*-styrene) block copolymer, but it would apply to any block copolymer with the structure A–B–A or (A–B)_{*n*}. In principle, A can be any polymer normally regarded as a hard thermoplastic, eg, polystyrene, poly(methyl methacrylate), or polypropylene, and B can be any polymer normally regarded as elastomeric, eg, polyisoprene, polybutadiene, polyisobutylene, or polydimethylsiloxane (Table 2).

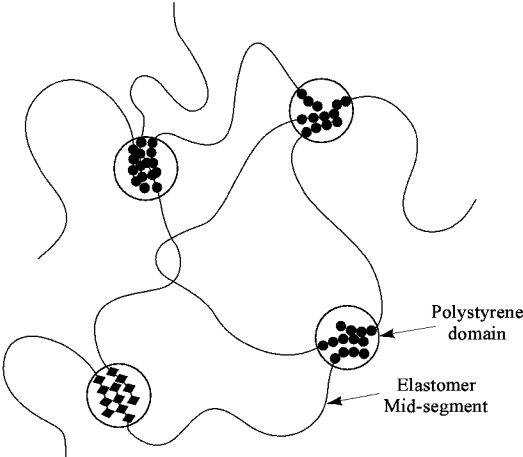


Fig. 1. Phase arrangement in styrenic block copolymers.

Table 2. Thermoplastic Block Copolymers

Hard segment, A	Soft or elastomeric segment, B	Structure	Refs.
polystyrene	polybutadiene, polyisoprene	A-B-A	4-6
poly(α -methylstyrene)	polybutadiene, polyisoprene	A-B-A	7
polystyrene	poly(ethylene- <i>co</i> -butylene)	A-B-A	5,6
polyethylene	poly(ethylene- <i>co</i> -butylene)	A-B-A	7
polystyrene	polydimethylsiloxane	A-B-A	8
poly(α -methylstyrene)	polydimethylsiloxane	A-B-A and (A-B) _n	7,8
polysulfone	polydimethylsiloxane	(A-B) _n	9
poly(silphenylene siloxane)	polydimethylsiloxane	(A-B) _n	10
polyurethane	polyester or polyether	(A-B) _n	11-13
polyester	polyether	(A-B) _n	14,15
polyamide	polyester or polyether	(A-B) _n	16-18
polycarbonate	polydimethylsiloxane	(A-B) _n	19-21
polycarbonate	polyether	(A-B) _n	22,23
polyetherimide	polydimethylsiloxane	(A-B) _n	24
polystyrene	polyisobutylene	A-B-A	25

Block copolymers with structures such as A-B or B-A-B are not thermoplastic elastomers, because for a continuous network to exist both ends of the elastomer segment must be immobilized in the hard domains. Instead, they are much weaker materials resembling conventional unvulcanized synthetic rubbers (4).

Five block copolymers are of commercial importance: poly(styrene-*b*-elastomer-*b*-styrene), thermoplastic polyurethanes, thermoplastic polyesters, thermoplastic polyamides, and polyetherimide–polysiloxane block copolymers. All but the first have the multiblock (A-B)_n structure. The morphology of the polyurethane, polyester, and polyamide block copolymers is shown diagrammatically in Figure 2. It has some similarities to that of poly(styrene-*b*-elastomer-*b*-styrene) equivalents (Fig. 1) and also some important differences: (1) the hard domains are more interconnected; (2) they are crystalline; and (3) these long (A-B)_n molecules may run through several hard and soft phases. The polyetherimide–polysiloxane block copolymers share some features of both types. As in the poly(styrene-*b*-elastomer-*b*-styrene) analogues, their hard domains are amorphous. However, these polymers have (A-B)_n structure rather than A-B-A, and so again, each molecule may run through several hard and soft phases.

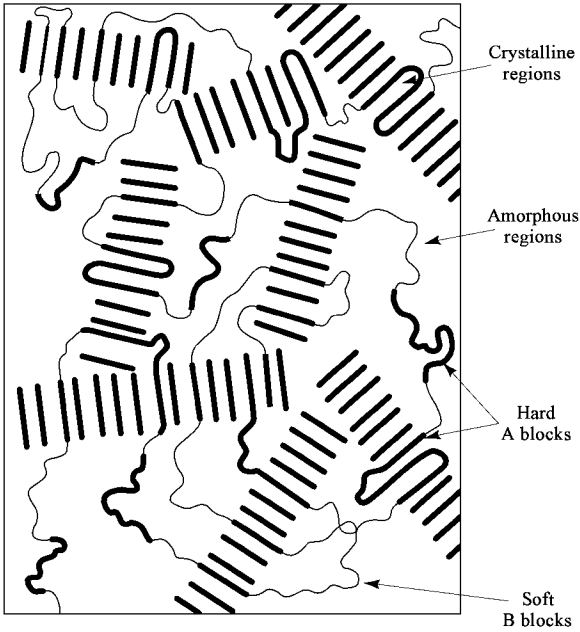


Fig. 2. Phase arrangement in crystalline block copolymers.

Not all thermoplastic elastomers are block copolymers. Those that are not are usually combinations of a hard thermoplastic with a softer, more rubberlike polymer. Usually, the components are mechanically mixed together, although it is sometimes possible to produce the rubber component *in situ* during polymerization (26). Typically, the two components form interdispersed multiphase systems and a diagrammatic structure for blends of this type is shown in Figure 3. Blends of polypropylene with ethylene–propylene–diene monomer (EPDM) were the first materials of this type (27,28). Blends with ethylene–propylene copolymer (EPR) are now more important commercially, and propylene copolymers often replace polypropylene homopolymer as the hard phase. However, some combinations of a hard thermoplastic with a rubberlike polymer are claimed to be single-phase systems (29). In some cases, the elastomer phase is cross-linked while the mixture is being highly sheared (30,31). This process is often referred to as "dynamic vulcanization" (32) and gives a finely dispersed and cross-linked elastomer phase (Fig. 4).

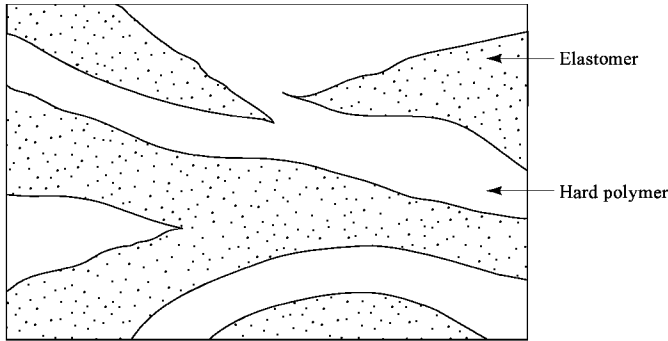


Fig. 3. Phase arrangement in hard polymer-elastomer blends.

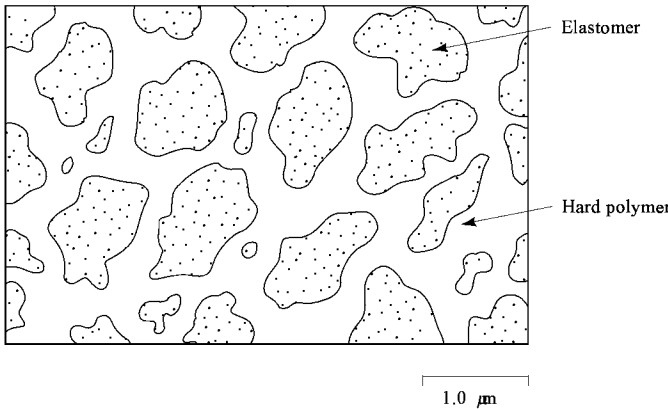


Fig. 4. Phase arrangement in hard polymer-elastomer combinations in which the elastomer phase has been dynamically vulcanized.

Other thermoplastic elastomer combinations, in which the elastomer phase may or may not be cross-linked, include blends of polypropylene with nitrile (30,31), butyl (33), and natural (34) rubbers, blends of PVC with nitrile rubber (35,36), and blends of halogenated polyolefins with ethylene interpolymers (29). Collectively, thermoplastic elastomers of this type are referred to herein as hard polymer-elastomer combinations. Some of the more important examples of the various types are shown in Table 3.

Table 3. Hard Polymer/Elastomer Combinations

Hard polymer	Elastomer	Structure ^a	Refs.
polypropylene	EPDM or EPR	B	27,28
polypropylene	EPDM	DV	30,31
polypropylene	butyl rubber	DV	33
polypropylene	natural rubber	DV	34
polypropylene	nitrile rubber	DV	30,31
PVC	nitrile rubber	B and DV	35,36
halogenated polyolefin	ethylene interpolymers	B	29

^a B = Blend of the two components (27); DV = dynamically vulcanized product in which the elastomer has been cross-linked during mixing (30).

Thermoplastic elastomers based on blends of a silicone rubber (cross-linked during processing) with block copolymer thermoplastic elastomers have also been described (37,38).

Property–Structure Relationships

Effects of variations in the structure of styrenic A–B–A block copolymers and similar materials were described in early work (4) and have been reviewed (5).

Molecular Weight. Compared with homopolymers of similar molecular weight, styrenic block copolymers have very high melt viscosities which increase with increasing molecular weight. These effects are attributed to the persistence of the two-phase domain structure in the melt and the extra energy required to disrupt this structure during flow. If the styrene content is held constant, the total molecular weight has little or no effect on the modulus of the material at ambient temperatures. This is attributed to the modulus of the elastomer phase being inversely proportional to the molecular weight between entanglements in the elastomer chains and the fact that this quantity is not affected by the total molecular weight.

Proportion of Hard Segments. As expected, the modulus of styrenic block copolymers increases with the proportion of the hard polystyrene segments. The tensile behavior of otherwise similar block copolymers with a wide range of polystyrene contents shows a family of stress–strain curves (4,7,8). As the styrene content is increased, the products change from very weak, soft, rubberlike materials to strong elastomers, then to leathery materials, and finally to hard glassy thermoplastics. The latter have been commercialized as clear, high impact polystyrenes under the trade name K-Resin (39) (Phillips Petroleum Co.). Other types of thermoplastic elastomers show similar behavior; that is, as the ratio of the hard to soft phase is increased, the product in turn becomes harder.

Elastomer Segment. The choice of elastomer segment has a profound effect on the properties of styrenic block copolymers. Four are commercially important: polybutadiene [9003-17-2], polyisoprene [9003-31-0], poly(ethylene-*co*-butylene) [9019-29-8], and poly(ethylene-*co*-propylene). The corresponding styrenic triblock copolymers are referred to as S–B–S, S–I–S, S–EB–S, and S–EP–S, respectively. Polybutadiene and polyisoprene both have one double bond per monomer unit. These double bonds are an obvious source of instability and limit the thermal and oxidative stability of the S–I–S and S–B–S block copolymers. In contrast, poly(ethylene-*co*-butylene) and poly(ethylene-*co*-propylene) are completely saturated, and so S–EB–S and S–EP–S block copolymers are much more stable. Another important aspect is the modulus of the materials. It is postulated that the modulus of these polymers is inversely proportional to the molecular weight between chain entanglements (M_e), as well as to the effects of the polystyrene domains acting as reinforcing filler particles (1,40). Values of M_e are as follows (41): polyisoprene (natural rubber), 6100; polybutadiene, 1900; and poly(ethylene-*co*-propylene),

1660. M_p for poly(ethylene-*co*-butylene) is similar to that of poly(ethylene-*co*-propylene) (42). Because of these differences in M_p , S–I–S block copolymers are softer than the S–B–S analogues; the S–EB–S and S–EP–S analogues are the hardest.

All these elastomers, especially poly(ethylene-*co*-butylene) and poly(ethylene-*co*-propylene), are nonpolar. The corresponding block copolymers can thus be compounded with hydrocarbon-based extending oils, but do not have much oil resistance. Conversely, block copolymers with polar polyester or polyether elastomer segments have little affinity for such hydrocarbon oils and so have better oil resistance.

Among the polyurethane, polyester, and polyamide thermoplastic elastomers, those with polyether-based elastomer segments have better hydrolytic stability and low temperature flexibility, whereas polyester-based analogues are tougher and have the best oil resistance (43). Polycaprolactones and aliphatic polycarbonates, two special types of polyesters, are used to produce premium-grade polyurethanes (12).

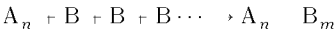
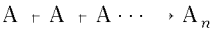
In the hard polymer elastomer combinations, the elastomer is often chosen to be a polar rubber or it is cross-linked; in some cases it is both. Either of these features improves the resistance to oils and solvents (44).

Hard Segment. The choice of the hard segment determines the upper service temperature and also influences the solvent resistance. In styrenic block copolymers, those based on poly(α -methylstyrene) [25014-31-7] have higher upper service temperature and tensile strength than analogues based on polystyrene [9003-53-6] (7); both are soluble in common solvents. Replacing the polystyrene end segments in S–EB–S by polyethylene (giving E–EB–E block copolymer) improves solvent resistance; the phases are not separated in the melt (7).

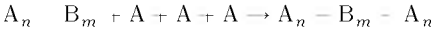
In thermoplastic polyurethanes, polyesters, and polyamides, the crystalline end segments, together with the polar center segments, impart good oil resistance and high upper service temperatures. The hard component in most hard polymer elastomer combinations is crystalline and imparts resistance to solvents and oils, as well as providing the products with relatively high upper service temperatures.

Synthesis

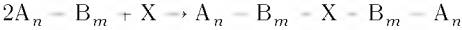
Block copolymers are synthesized by a variety of methods (45,46); most important are sequential polymerization and step growth. In sequential polymerization, a polymer (A_n) is first synthesized in such a way that it contains at least one group per molecule that can initiate polymerization of another monomer B.



The product can be converted to a triblock copolymer by further addition of A:



In a variation of this process, polymerization can start in the center (B_m) segment, and A_n segments can then be polymerized onto each end. Alternatively, $A_n - B_m$ can be joined together by a coupling agent:

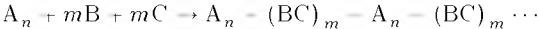


In this example, X is difunctional and the product is linear. If the functionality of X is higher, the product is branched, ie, it is a star polymer.

In step-growth polymerization, one or both segments can be produced separately as difunctional prepolymers. The products can then be linked together:



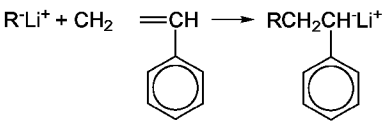
or react with more difunctional monomer:



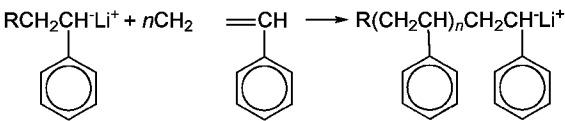
Thermoplastic elastomers that are hard polymer elastomer combinations are often not truly synthesized. Instead, the two polymers that form the hard and soft phases are intimately mixed on high shear equipment.

Commercial Production

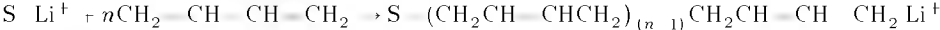
Commercially, the poly(styrene-*b*-elastomer-*b*-styrene) materials are made by anionic polymerization (7,45–47). An alkyllithium initiator (RLi) first reacts with styrene [100-42-5] monomer:



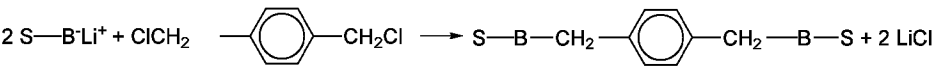
This product acts as an initiator for further polymerization:



The product, referred to here as S[•]Li⁺, is able to initiate further polymerization. Similar products have been termed living polymers (48). Addition of a second monomer, such as butadiene [106-99-0], gives



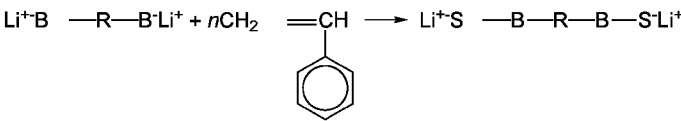
The product of this reaction (S–B[•]Li⁺) may initiate a further reaction with styrene monomer to give S–B–S[•]Li⁺. This, in turn, can react with an alcohol, ROH, to give S–B–SH + LiOR. Alternatively, S–B[•]Li⁺ may react with a coupling agent such as an organohalogen (45):



Other coupling agents include esters (45), chlorosilanes (45,46), and divinylbenzene (46). The last gives highly branched materials, whereas the others can give branched or linear products, depending on the functionality of the coupling agent. Another variation of anionic polymerization uses multifunctional initiators (48), in which the polymer chains grow outward from the center of the molecule. In the case considered here, a difunctional initiator (LiRLi) first reacts with butadiene monomer:



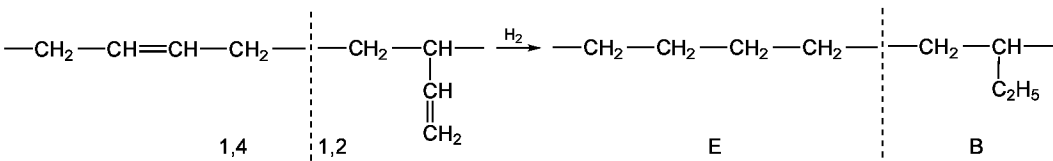
This product is a difunctional initiator and can polymerize styrene monomer:



and the final product can again react with an alcohol to give S–B–S, ignoring the minor amount of R moiety at the center of the molecule.

All these polymerizations proceed only in the absence of oxygen or water, which react with the highly reactive propagating species. Polymerization is usually carried out in an inert, hydrocarbon solvent and under a nitrogen blanket. Under these conditions, polymers with narrow molecular-weight distributions and precise molecular weights can be produced in stoichiometric amounts.

Commercially, anionic polymerization is limited to three monomers: styrene, butadiene, and isoprene [78-79-5]; therefore only two useful A–B–A block copolymers, S–B–S and S–I–S, can be produced directly. In both cases, the elastomer segments contain double bonds which are reactive and limit the stability of the product. To improve stability, the polybutadiene mid-segment can be polymerized as a random mixture of two structural forms, the 1,4 and 1,2 isomers, by addition of an inert polar material to the polymerization solvent; ethers and amines have been suggested for this purpose (46). Upon hydrogenation, these isomers give a copolymer of ethylene and butylene.

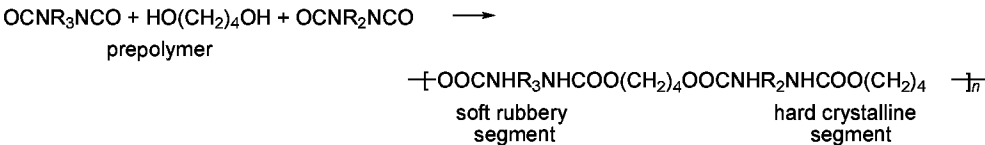


The S–EB–S block copolymers produced in this way have excellent resistance to degradation (49). Similarly, S–I–S block copolymers can be hydrogenated to give the more stable S–EP–S equivalents.

Thermoplastic polyurethane elastomers are produced from prepolymers by polycondensation (12,13). A relatively high molecular-weight polyester or polyether with terminal hydroxy groups (a polyglycol) first reacts with an excess of a diisocyanate.



Depending on the excess diisocyanate, this product reacts further with more polyglycol to give a prepolymer in which the segments originally derived from the polyether or polyester glycol and the urethane alternate, and which is terminated by isocyanate groups. This prepolymer, designated OCNR₃NCO, becomes the elastomeric or soft segment in the final polymer. It, in turn, reacts with a low molecular-weight glycol such as butanediol and with more diisocyanate.



The segments derived from the condensation reaction of the butanediol and the diisocyanate agglomerate into separate phases, which are hard and crystalline. The elastomeric chains are thus cross-linked to form a network similar in many ways to that given by the simple poly(styrene-*b*-elastomer-*b*-styrene) polymer described previously (4,5,11). However, the polymer is a multiblock (A–B)_{*n*} rather than a simple A–B–A triblock, and the molecular weights and molecular-weight distributions of the segments are not as well controlled. 4,4′-Diphenylmethane diisocyanate is the most common diisocyanate used in this application (13) but toluene and *para*-phenylene (50) diisocyanates are alternatives.

Commercial thermoplastic polyesters are synthesized in a similar way by the reaction of a relatively high molecular-weight polyether glycol with butanediol and dimethyl terephthalate (14,15). The polyether chain becomes the soft segment in the final product, whereas the terephthalic acid–butanediol copolymer forms the hard crystalline domains.

There are several types of thermoplastic polyamides (18). Their synthesis is similar to that of the polyurethane and polyester equivalents (16–18). In some cases, only the soft segments are prepolymers; whereas in others, prepolymers are used to give both segmental types. Both polyesters and polyethers are used for the soft segments, and this choice affects the final properties of the product. Various polyamides, including those based on aromatic groups, may be used for the hard segments. These all have different melting points and degrees of crystallinity and so give products with a wide range of properties.

Information on the synthesis of the polyetherimide–polysiloxane block copolymers has not been disclosed. Many other synthetic methods for preparing block copolymers have been described (19,20,25) but are currently not believed to be commercially important.

The production of the hard polymer–elastomer combinations is more simple. The two components are mixed together under conditions of intensive shear. In some cases, grafting may occur. In a variation of this technique, the elastomer can be cross-linked while the mixing is taking place, a process described as dynamic vulcanization (32).

A large number of hard polymer–elastomer combinations made by the last technique have been investigated (30). In some cases, the components are technologically compatibilized by use of a grafting reaction, but usually a fine dispersion of the two phases is formed that is sufficient to give the product the properties of a thermoplastic elastomer.

Economic Aspects

Global consumption of thermoplastic rubbers of all types is estimated at about 600,000 t/yr (51). Of this, 42% was estimated to be consumed in the United States, 39% in Western Europe, and 19% in Japan. At present, the worldwide market is estimated to be divided as follows: styrenic block copolymers, 48%; hard polymer–elastomer combinations, 26%; thermoplastic polyurethanes, 12%; thermoplastic polyesters, 4%; and others, 9%. The three largest end uses were transportation, 23%; footwear, 18%; and adhesives, coatings, etc, 16%. The ranges of the hardness values, prices, and specific gravities of commercially available materials are given in Table 4.

Table 4. Price and Property Ranges for Thermoplastic Elastomers^a

Elastomer	Approximate price range, \$/kg	Specific gravity	Hardness range ^b
polystyrene–elastomer block copolymers			

S–B–S (pure)	1.9–2.6	0.94	65A–75A
S–I–S (pure)	2.2–2.9	0.92	32A–37A
S–EB–S (pure)	4.1–5.9	0.91	65A–75A
S–B–S (compounds)	1.3–3.1	0.9–1.1	40A–45D
S–EB–S (compounds)	2.4–4.4	0.9–1.2	18A–60D
polyurethane elastomer block copolymers	4.4–7.2	1.05–1.25	70A–75D
polyester elastomer block copolymers	4.6–7.7	1.15–1.40	35D–70D
polyamide elastomer block copolymers	7.5–8.8	1.0–1.15	60A–65D
polyetherimide polysiloxane block copolymers	37	1.18	70D
polypropylene EPDM or EPR blends	2.0–3.3	0.9–1.05	60A–65D
polypropylene EPDM dynamic vulcanizates	3.6–5.5	0.95–1.0	45A–50D
polypropylene butyl rubber dynamic vulcanizates	4.6–7.9	0.95–1.0	40A–80A
polypropylene natural rubber dynamic vulcanizates	2.6–3.5	1.0–1.05	60A–50D
polypropylene nitrile rubber dynamic vulcanizates	4.4–5.5	1.0–1.1	70A–50D
PVC nitrile rubber blends	2.9–3.1	1.20–1.33	50A–80A
halogenated polyolefin ethylene interpolymer blends	3.5–4.4	1.10–1.25	60A–80A

^a These price and property ranges do not include fire retardant grades or highly filled materials for sound deadening.

^b Shore A or D as indicated.

Applications

Trade names and suppliers of commercial thermoplastic elastomers of all types are given in Tables 5–7.

Table 5. Trade Names of Thermoplastic Elastomers Based on Styrenic Block Copolymers^a

Trade name	Manufacturer	Type	Elastomer segment	Notes
<i>General-purpose, soluble^b</i>				
Kraton D and Cariflex	Shell	linear and branched	B or I	also compounded products
Vector ^c	Dexco	linear	B or I	
Solprene ^d	Phillips	branched	B	
Finaprene	Fina	linear	B	improved stability only compounded products wire and cable compounds only compounded products elastomer block is hydrogenated polyisoprene medical applications; contains silicone oil high polystyrene content very high polystyrene content; hard and rigid
Tufprene and Asaprene	Asahi	linear	B	
Europrene Sol T	Enichem	linear and branched	B or I	
Quintac	Nippon Zeon	linear	I	
Kraton G	Shell	linear	EB or EP	
Dynaflex	GLS ^e	linear	B or EB	
Elexar	Shell ^f	linear	EB	
Flexprene	Teknor Apex	linear	B	
Tekron	Teknor Apex	linear	EB	
Septon	Kuraray	linear	EP	
C-Flex	Concept	linear	EB	<i>Specialty materials</i>
Stereon	Firestone	linear	B	
K-Resin	Phillips	branched	B	

^a Hard segment styrene.

^b Not available as compounded products unless otherwise noted.

^c Joint venture of Dow and Exxon.

^d No longer made in the United States.

^e Great Lakes Terminal and Transport Corp.

^f Now produced by Teknor Apex Co.

Table 6. Trade Names of Multiblock Thermoplastic Elastomers Based on Polyurethane/Elastomer, Polyether/Elastomer, and Polyamide/Elastomer Block Copolymers

Trade name	Manufacturer	Hard segment	Elastomer segment	Notes
Estane	B. F. Goodrich	polyurethane	polyether or polyester (amorphous)	hard, abrasion- and oil-resistant; high cost
Q-Thane ^a	Morton International			
Pellethane ^a	Dow			
Texin	Miles ^b			
Elastollan	BASF	polyester	polyether	similar to polyurethane and better at low temperatures
Urafil	Akzo			
Hytel	Du Pont			
Lomod	GE			
Ecdel	Eastman			

Riteflex	Hoechst-Celanese			
Pebax	Atochem			
Vestenamer	HbIs			
Grilamid	Emser	polyamide		
Montac ^c	Monsanto		polyether or polyester (amorphous)	
Orevac ^c	Atochem			similar to polyurethanes but can be softer; very good at low temperatures
Siltem	GE	polyether imide	polysiloxane	fire retardant; used in wire and cable insulation

^a Including some with polycaprolactone segments.

^b Formerly Mobay. Includes some polymers with aliphatic polycarbonate segments (trade named: Desmopan).

^c For hot-melt adhesives.

Table 7. Some Trade Names of Thermoplastic Elastomers Based on Hard Polymer/Elastomer Combinations

Trade name	Manufacturer	Type	Elastomer	Notes
<i>Polypropylene hard polymer</i>				
TPR	AES ^a			
Ren-Flex	Dexter			
Polytrope	Schulman			
Telcar	Teknor Apex	blend	EPDM or EPR	relatively hard, low density, not highly filled
Ferroflex	Ferro	}		
Hifax	Himont			
Santoprene	AES			
Sarlink 3000	Novacor	DV ^b	EPDM	
Telprene	Teknor Apex			better oil resistance, low compression set, softer
Hifax XL	Himont	}		
Trefsin	AES			
Sarlink 2000	Novacor	DV	butyl rubber	
Vyram	AES	DV	natural rubber	low permeability, high damping
Geolast	AES	DV	nitrile rubber	low cost
<i>Poly(vinyl chloride) hard polymer</i>				
Sarlink 1000	Novacor	DV		
Chemigum	Goodyear	blend	nitrile	
Ultraprene	Teknor Apex	blend	rubber	oil-resistant
Elastar	Nippon Zeon	^c		
<i>Other hard polymers</i>				
Alcryn	Du Pont	blend ^d	ethylene interpolmer	single-phase, soft, oil-resistant
Rimplast	Petrarch Systems	blend ^e		medical applications

^a Advanced Elastomer Systems; a joint venture between Monsanto and Exxon.

^b The soft phase is dynamically vulcanized, ie, cross-linked during mixing (32).

^c Elastomer phase reported to be ionically cross-linked (52).

^d Chlorinated polyolefin hard polymer.

^e Blend of other thermoplastic elastomers with silicone rubbers.

Styrenic Block Copolymers. The applications of these block copolymers are described in detail in Reference 6.

Substitute for Conventional Vulcanized Rubbers. For this application, the products are processed by techniques and equipment developed for conventional thermoplastics, ie, injection molding, extrusion, etc. The S–B–S and S–EB–S polymers are preferred (small amounts of S–EP–S are also used). To obtain a satisfactory balance of properties, they must be compounded with oils, fillers, or other polymers; compounding reduces costs. Compounding ingredients and their effects on properties are given in Table 8. Oils with high aromatic content should be avoided because they plasticize the polystyrene domains. Polystyrene is often used as an ingredient in S–B–S-based compounds; it makes the products harder and improves their processibility. In S–EB–S-based compounds, crystalline polyolefins such as polypropylene and polyethylene are preferred. Some work has been reported on blends of liquid polysiloxanes with S–EB–S block copolymers. The products are primarily intended for medical and pharmaceutical-type applications and hardnesses as low as 5 on the Shore A scale have been reported (53).

Table 8. Compounding Styrenic Block Copolymers

Properties	Component				
	Oils	Polystyrene	Polyethylene	Polypropylene	Fillers
hardness	decreases	increases	increases	increases	small increase
processibility	improves	improves	improves	improves, especially with S–EB–S	improves
effect on oil	none	none	improves	improves	none

resistance					
cost	decreases	decreases	decreases	decreases	decreases
other	decreases uv		often gives satin	improves high	often improves surface
	resistance		finish	temperature properties	appearance

Large amounts of inert fillers, such as whiting, talc, and clays, can be added. Very dense fillers, such as barium or strontium sulfates, are used to make compounds intended for sound-deadening applications. In contrast, high levels of reinforcing fillers, such as carbon black, produce undesirable properties in the final product.

A large volume usage of S–B–S-based compounds is in footwear. Canvas footwear, such as sneakers and unit soles, can be made by injection molding. Frictional properties resemble those of conventionally vulcanized rubbers and are superior to those of the flexible thermoplastics, such as plasticized poly(vinyl chloride). The products remain flexible under cold conditions because of the good low temperature properties of the polybutadiene segment.

Compounds based on S–EB–S usually contain polypropylene, which improves solvent resistance and processibility and raises upper service temperatures. Compounds intended for use in the automotive industry are able to survive 1000 hours air exposure at temperatures of 125°C with only minor changes in properties (54). Very soft compounds have been developed to replace foam rubber for interior trim parts. In this and similar applications, these soft compounds are usually insert molded over polypropylene or metal and then coated with flexible polyurethane paint (55). Other automotive applications include products intended for sound deadening, flexible air ducts, and gear shifter boots, as well as improving the properties of sheet molding compounds.

Other uses for which special compounds have been developed include materials intended for food contact, wire insulation, and pharmaceutical applications.

Commercial products have hardnesses from 28 on the Shore A scale (which is very soft) to 45 on the Shore D scale (almost leathery). Specific gravities usually range from 0.9 to 1.20; some products intended for soundproofing have specific gravities as high as 1.95. Properties of representative grades are given in Table 9.

Table 9. Properties of Compounded Styrenic Block Copolymers

Product	Kraton D2109	Kraton D3202 ^a	Dynaflax D3202 ^a	Kraton D5119	Kraton D5298	Kraton G7535X	Kraton G7528X	Kraton G7702	Kraton G7715	Kraton G7720	Elexar 8431	Elexar 8451
application	milk tubing	medical	general-purpose	footwear unitsole	footwear slipper-sole	soft function-alized	automotive very soft	automotive low compression set	automotive general-purpose	automotive general-purpose	wire and cable	wire and cable
hardness, Shore A or D	48A ^b	55A	50A	55A	47A	35A	28A	36A	57A	60A	64A	82A ^b 34D
tensile strength, MPa ^c	11 ^b	6.0	5.6	3.8	3.9	7.0	5.8	4.6	12 ^b	5.0	9.6 ^b	12 ^b
300° modulus, MPa ^c	2.0 ^b	2.8	4.3	3.5	1.9	1.3	1.0	3.2	2.6 ^b	2.5	3.6 ^b	4.4 ^b
elongation, %	850 ^b	700	400	360	550	900	950	650	750 ^b	700	650 ^b	650 ^b
specific gravity	0.94	0.90	1.0	1.09	0.96	0.92	0.91	1.07	0.91	1.19	0.92	1.01

^a Formerly Kraton D3202.

^b Measured on extruded samples, all others measured on injection molded samples.

^c To convert MPa to psi, multiply by 145.

Processing is relatively easy. In general, products based on S–B–S are processed under conditions appropriate for polystyrene, whereas products based on S–EB–S are processed under conditions appropriate for polypropylene. Pre-drying is not needed and scrap is recycled.

Adhesives, Coatings, and Sealants. For these applications, styrenic block copolymers must be compounded with resins and oils (Table 10) to obtain the desired properties (56–58). Materials compatible with the elastomer segments soften the final product and give tack, whereas materials compatible with the polystyrene segments impart hardness. The latter are usually styrenic resins with relatively high softening points. Materials with low softening points are to be avoided, as are aromatic oils, since they plasticize the polystyrene domains and reduce the upper service temperature of the final products.

Table 10. Resins Used to Formulate Adhesives, Sealants, Etc From Styrenic Block Copolymers

Resin type	Segment compatibility ^a
polymerized C ₅ resins (synthetic polyterpenes)	I
hydrogenated rosin esters	B
saturated hydrocarbon resins	EB
naphthenic oils	I, B
paraffinic oils	EB
low molecular-weight polybutenes	EB
aromatic resins	S

^a I indicates compatible with polyisoprene segments; B, compatible with polybutadiene segments; EB, compatible with poly(ethylene–butylene) segments; and S, compatible with poly-styrene segments.

These resins and oils have low molecular weights, ie, typically below 1000. This, combined with the relatively low molecular weights of the styrenic

block copolymers, ie, 30,000–150,000, allows solutions in common solvents to be formulated at high solids levels. Alternatively, the products can be applied as hot melts, with considerable advantages in terms of safety, production rates, energy consumption, and air pollution.

Blends with Thermoplastics or Other Polymeric Materials. Styrenic block copolymers are technologically compatible with a surprisingly wide range of materials and can be blended to give useful products (59). Blending can often be carried out on the equipment producing the final article. Blends of S–B–S with polystyrene, polyethylene, or polypropylene show improved impact and tear resistance. Similarly, S–EB–S can be blended with the less polar engineering thermoplastics such as poly(phenylene oxide) and polycarbonate. An unusual feature of these block copolymers is their ability to enable useful blends to be made from incompatible polymers, eg, polystyrene or poly(butylene terephthalate) with polyethylene (60). A new development is the use of functionalized S–EB–S block copolymers as impact modifiers for more polar engineering thermoplastics such as polyesters and polyamides. The functionality is given by maleic acid anhydride groups grafted to the S–EB–S polymer chain. These functionalized S–EB–S block copolymers have also been found useful in the compatibilization of polyolefins with polyamides (61) and with poly(phenylene ether) (62).

Special grades of styrenic block copolymers are useful modifiers for sheet molding compounds (SMC) based on thermoset polyesters. They improve surface appearance, impact resistance, and hot strength.

Blends with styrenic block copolymers improve the flexibility of bitumens and asphalts. The block copolymer content of these blends is usually less than 20°%; even as little as 3°% can make significant differences to the properties of asphalt (qv). The block copolymers make the products more flexible, especially at low temperatures, and increase their softening point. They generally decrease the penetration and reduce the tendency to flow at high service temperatures; and they also increase the stiffness, tensile strength, ductility, and elastic recovery of the final products. Melt viscosities at processing temperatures remain relatively low so the materials are still easy to apply. As the polymer concentration is increased to about 5°%, an interconnected polymer network is formed. At this point the nature of the mixture changes from an asphalt modified by a polymer to a polymer extended with an asphalt.

It is important to choose the correct grade of asphalt; those with a low asphaltene content and or high aromaticity in the maltene fraction usually give the best results (63,64). Applications include road surface dressings such as chip seals (applied to hold the aggregate in place when a road is resurfaced); slurry seals; hot-mix asphalt concrete (a mixture of asphalt and aggregate used in road surfaces); road crack sealants; roofing; and other waterproofing and adhesive applications (64–66). Because of their lower cost, S–B–S block copolymers are usually chosen for this application; but in roofing and paving applications, the S–EB–S block copolymers are also used because of better long-term aging resistance.

Multiblock Copolymers. Replacement of conventional vulcanized rubber is the main application for the polar polyurethane, polyester, and polyamide block copolymers. Like styrenic block copolymers, they can be molded or extruded using equipment designed for processing thermoplastics. Melt temperatures during processing are between 175 and 225°C, and predrying is required; scrap is reusable. They are mostly used as essentially pure materials, although some work on blends with various thermoplastics such as plasticized and unplasticized PVC and also ABS and polycarbonate (14,18,67–69) has been reported. Plasticizers intended for use with PVC have also been blended with polyester block copolymers (67).

All three types of these block copolymers are relatively hard (from 70 on the Shore A scale to 70 on the Shore D scale) with specific gravities of 1.00–1.25. Properties of individual grades have been described in detail (Tables 11–13). Applications (13,15,18) taking advantage of toughness, abrasion resistance, flexibility, and resistance to oils and solvents include belting, hydraulic hose, tires, shoe soles, wire coatings, and automobile parts. The surface of these molded parts can be easily painted or metallized. Medical uses, such as implants, are a significant application for the polyurethane block copolymers with polyether elastomer segments (70). Blends with silicone rubbers have been described for these applications (37,38). Another application is the use of a clear polyester thermoplastic elastomer as a replacement for glass bottles in medical applications (3).

Table 11. Properties of Polyurethane/Elastomer Block Copolymers

Product	Estane 58133	Estane 58311	Q-Thane PC86	Texin 480A	Pellethane 2102-90A	Pellethane 2103-70A
soft segment type	polyester	polyether	polyester	polyester	polycapro-lactone	polyether
hardness, Shore A or D	55D	85A	50D	86A	90A	70A
tensile strength, MPa ^a	40	47	52	41	48	24
100°% modulus, MPa ^a	14	6.2	8.3	5.1	10.7	3.5
elongation, °%	500	530	475	500	500	500
specific gravity	1.22	1.11		1.20	1.20	1.06

^a To convert MPa to psi, multiply by 145.

Table 12. Properties of Polyester/Elastomer Block Copolymers

Product	Hytrel 4056	Hytrel 5526	Hytrel 6356	Hytrel 7246	Lomod ST3090A	Lomod TE3055A
hardness, Shore A or D	40D, 92A	55D	63D	72D	90A 31D	54D
tensile strength, MPa ^a	26	40	41	46	14	23
100°% modulus, MPa ^a	9.5	19	20 ^b	25 ^b		
elongation, °%	550	500	420	350	250	290
specific gravity	1.17	1.20	1.22	1.25	1.13	1.22

^a To convert MPa to psi, multiply by 145.

^b After yielding at lower elongation.

Table 13. Typical Properties of Polyamide/Elastomer and Polyetherimide/Elastomer Block Copolymers

Product	Pebax 2533	Pebax 4033	Pebas 6333	Grilamid Ely 60	Siltem ^a STM1500
hardness, Shore A or D	25D 75A	40D	63D	62D	69D
tensile strength, MPa ^b	29	33	49	36	25
100°% modulus, MPa ^b	4.3	10	19		
elongation, °%	350	620	680	300	105
specific gravity	1.01	1.01	1.01	1.01	1.18

^a Fire retardant grade; intended for wire insulation.

^b To convert MPa to psi, multiply by 145.

Some grades of polyurethane and polyester copolymers are used as hot-melt adhesives. Applications include shoe manufacture and as an adhesive interlayer in coextrusion.

The polyetherimide–polysiloxane multiblock copolymers are relatively hard (about 70 on the Shore D scale). Their main application is flame-resistant wire and cable covering (24), where they combine very low flammability with a low level of toxic products in the smoke. This unusual and vital combination of properties justifies their relatively high price, about \$37 kg, at a specific gravity of about 1.2.

Hard Polymer/Elastomer Combinations. Substitution of conventional vulcanized rubbers is the main application for these materials (27,28,31). The first ones to be developed were those based on polypropylene and EPDM. To begin with, simple mechanical mixtures of the two components were introduced. These had some limitation, because EPDM in its unvulcanized state has almost no tensile strength or oil resistance. Thus only fairly hard products, ie, those containing small amounts of EPDM, had satisfactory properties and so the first products of this type had hardness values in the Shore D range, although softer versions are now available. In later work, products based on dynamic vulcanization (32) were produced. In this process, the EPDM is simultaneously mixed and cross-linked. The result is a very fine dispersion of vulcanized EPDM particles in a polypropylene matrix. The improved properties of the EPDM phase allow much higher levels to be used, giving quite soft products (as low as about 45 on the Shore A scale). Compared to the mixtures of EPDM with polypropylene, the dynamic vulcanizates have lower compression set and better oil resistance. Natural and butyl rubbers have been used to replace EPDM in similar dynamically vulcanized products (33,34). Those based on natural rubber are low in cost and have properties intermediate between the EPDM-based dynamic vulcanizates and the simple EPDM-based mixtures. Those based on butyl rubber have low gas permeability and high damping, thus they are often used as vapor barriers or vibration isolators. An unexpected property advantage with those based on butyl rubber is that some grades show excellent adhesion when insert molded against such polar engineering thermoplastics as poly(butylene terephthalate) and polyamides (71). Hardness values are between about 60 and 80 on the Shore A scale. Specific gravities of these various combinations of polypropylene with EPDM, butyl, and natural rubbers are between about 0.9 and 1.05. Properties of representative grades are given in Table 14. Molding and extrusion conditions are similar to those used for polypropylene, and the scrap is reusable. Important applications are wire insulation, appliance parts, and automobile exterior and interior parts (both painted and unpainted).

Table 14. Properties of Hard Polymer/Elastomer Combinations

Product	Vistaflex		Republic AB	Telcar	Santoprene		Trefsin
	9101-65-W90	9013-45-W90			101-73	103-40	
	0	0	6053 ^a	302	201-73	203-40	3201-60
hardness, Shore A or D	65A	45D	45D	75A	73A	40D	60A
tensile strength, MPa ^b	4.5	15	4.8	11	8.3	19	4.6
100° modulus, MPa ^b	2.2	11			3.2	8.6	1.9
elongation, %	525	750	350	900	375	600	355
specific gravity	1.00	0.95	1.85	0.88	0.98	0.95	0.97

^a Grade intended for sound-deadening applications.

^b To convert MPa to psi, multiply by 145.

Even though vulcanized EPDM has some oil resistance, in contrast to unvulcanized EPDM which has virtually none, a rubber with inherent oil resistance should be even better. For this reason, more polar rubbers have replaced EPDM in applications where oil resistance is critical. Dynamic vulcanizates of nitrile rubber with polypropylene are one example. Commercial grades are somewhat harder than EPDM equivalents (between 70 on the Shore A scale to 40 on the Shore D scale) and are also more dense (about 1.0 to 1.1 specific gravity). PVC nitrile blends and dynamic vulcanizates also have excellent oil resistance as well as resistance to flex cracking and abrasion (36). Hardness ranges from about 50 to 80 on the Shore A scale and specific gravity from about 1.2 to 1.3. Blends of halogenated polyolefins with ethylene interpolymers are claimed to give single-phase systems (29). They have a very rubberlike feel and are similar to the PVC nitrile rubber combinations as far as hardness and specific gravity are concerned. Like them, they have good oil resistance. All these products based on the more polar rubbers are used in molded appliance, automotive, and similar rubberlike parts where oil resistance is needed at a reasonable cost. Properties of representative grades are given in Table 15.

Table 15. Properties of Hard Polymer/Elastomer Combinations Based on Polar Elastomers

Product	Geolast		Alcryn MPR			Sarlinc	Chemigum		
	701-70	701-87	60A	70A	80A		03050	031	03070
						1160	50		70
hardness, Shore A or D	70A	87A	62A	69A	78A	57A	50A		70A
tensile strength, MPa ^a	6.2	12	12	13	13	11	12		17
100° modulus, MPa ^a	3.3	6.0	3.7	4.5	7.2	3.4	2.9		5.9
elongation, %	265	380	325	295	210	345	470		400
specific gravity	1.00	0.98	1.21	1.23	1.25	1.17	1.20		1.23

^a To convert MPa to psi, multiply by 145.

Health and Safety

Most thermoplastic elastomers are stable materials and decompose only slowly under normal processing conditions. If decomposition does occur, the products are usually not particularly hazardous and should not present a problem if good ventilation is provided. Extra caution should be exercised when processing polyurethanes, especially those containing polycaprolactone segments. In these cases the decomposition products may include isocyanates and caprolactam, both of which are potential carcinogens.

Of course, all materials that are processed in the molten state can cause burns if the hot material comes in contact with the skin. Care must be taken to avoid this, and it should be noted that molten material left in the barrel of an extruder or injection molding machine can "spit" unexpectedly. In all cases, it is recommended that the manufacturer's Material Safety Data Sheet be consulted before working with any of these materials.

Reprocessing

Easy reprocessing is one of the great advantages that thermoplastic elastomers have over conventional vulcanized rubbers. The scrap can be reground and is usually blended with virgin material before being reworked. Regrinding is not difficult if it is remembered that rubber must be cut rather than shattered. This means that the cutter blades must be sharp and clearances minimized. It is usually best to dry the ground scrap before reworking it, and for the polyurethanes, polyesters, and polyamides, drying is a necessity. Thermoplastic elastomers can also be used to "sweeten" regrind; that is, they can be

blended with reground scrap from conventional thermoplastics to restore impact strength and reduce brittleness. Many applications, eg, coextrusion, generate mixed scrap, which usually has very poor properties. Thermoplastic elastomers can often convert this into useful material (60–62).

BIBLIOGRAPHY

"Elastomers, Synthetic (Thermoplastic)" in *ECT* 3rd ed., Vol. 8, pp. 626–640, by A. F. Finelli, R. A. Marshall, and D. A. Chung, The Goodyear Tire & Rubber Co.

1. N. R. Legge, G. Holden, and H. E. Schroeder, eds., *Thermoplastic Elastomers—A Comprehensive Review*, Carl Hanser Verlag and Oxford University Press, Munich, New York, 1987. This reference and the next are recommended for readers who wish to obtain more detailed information on the subject.
2. B. M. Walker and C. P. Rader, eds., *Handbook of Thermoplastic Elastomers*, 2nd ed., Van Nostrand Reinhold, New York, 1988.
3. R. School, in Ref. 2, Chapt. 9.
4. G. Holden, E. T. Bishop, and N. R. Legge, *J. Poly. Sci. C*(26), 37 (1969).
5. G. Holden and N. R. Legge, in Ref. 1, Chapt. 3.
6. W. M. Halper and G. Holden, in Ref. 2, Chapt. 2.
7. M. Morton, in Ref. 1, Chapt. 4.
8. J. C. Saam, A. Howard, and F. W. G. Fearon, *J. Inst. Rubber Ind.* **7**, 69 (1973).
9. A. Noshay, M. Matzner, and C. N. Merriam, *J. Poly. Sci. A-1* (9), 3147 (1971).
10. R. L. Merker, M. J. Scott, and G. G. Haberland, *J. Poly. Sci., A* **2**, 31 (1964).
11. S. L. Cooper and A. V. Tobolsky, *Tex. Res. J.* **36**, 800 (1966).
12. W. Merkel, W. Goyert, and W. Wieder, in Ref. 1, Chapt. 2.
13. E. C. Ma, in Ref. 2, Chapt. 7.
14. R. K. Adams and G. K. Hoeschele, in Ref. 1, Chapt. 8.
15. T. W. Sheridan, in Ref. 2, Chapt. 6.
16. R. G. Nelb, A. T. Chen, and K. Onder, in Ref. 1, Chapt. 9A.
17. G. Deleens, in Ref. 1, Chapt. 9B.
18. W. J. Farrissey and T. M. Shah, in Ref. 2, Chapt. 8.
19. H. A. Vaughn, *J. Poly. Sci. B* **7**, 569 (1969).
20. R. P. Kambour, *J. Poly. Sci. B* **7**, 573 (1969).
21. D. G. LeGrand, *J. Poly. Sci. B* **7**, 579 (1969).
22. E. P. Goldberg, *J. Poly. Sci. C* **4**, 707 (1963).
23. K. P. Perry, W. J. Jackson, Jr., and J. R. Caldwell, *J. Appl. Poly. Sci.* **9**, 3451 (1965).
24. J. Mihalich, paper presented at the *2nd International Conference on Thermoplastic Elastomer Markets and Products*, Schotland Business Research, Orlando, Fla., Mar. 15–17, 1989.
25. G. Kaszas, J. E. Puskas, W. G. Hager, and J. P. Kennedy, *J. Polym. Sci., Polym. Chem. Ed.*, in press.
26. C.-K Shih and A. C. L. Su, in Ref. 1, Chapt. 5.
27. J. R. Wolfe, Jr., in Ref. 1, Chapt. 6.
28. C. D. Shedd, in Ref. 2, Chapt. 3.
29. J. G. Wallace, in Ref. 2, Chapt. 5.
30. A. Y. Coran, in Ref. 1, Chapt. 7.
31. C. P. Rader, in Ref. 2, Chapt. 4.
32. **U.S. Pat. 3,037,954** (June 5, 1962), A. M. Gessler (to Esso Research & Engineering Co.).
33. R. C. Puydak, in Ref. 24.
34. A. J. Tinker, paper presented at the *Symposium on Thermoplastic Elastomers*, ACS Rubber Division, Cincinnati, Ohio, Oct. 18–21, 1988.
35. M. Stockdale, in Ref. 34.
36. P. Tandon and M. Stockdale, paper presented at the *4th International Conference on Thermoplastic Elastomer Markets and Products*, Schotland Business Research, Orlando, Fla., Feb. 13–15, 1991.
37. **U.S. Pat. 4,500,688** (Feb. 19, 1985), B. C. Arkles (to Petrarch Systems).
38. B. C. Arkles, *Med. Dev. Diagnost. Ind.* **5**(11), 66 (1983).
39. Technical bulletin, TIB 200, Phillips 66 Co., 1991.
40. J. K. Bard and C. I. Chung, in Ref. 1, Chapt. 12, Section 1.
41. J. D. Ferry, *Viscoelastic Properties of Polymers*, John Wiley & Sons, Inc., New York, 1980, p. 374.
42. L. J. Fetters, private communication, 1983.
43. Ref. 13, pp. 233, 237, and 238.
44. Ref. 31, p. 86.
45. N. R. Legge, S. Davison, H. E. De La Mare, G. Holden, and M. K. Martin in R. W. Tess and G. W. Poehlein, eds., *Applied Polymer Science*, American Chemical Society, Washington, D. C., 1975, Chapt. 9, pp. 189–200.
46. P. Dreyfuss, L. J. Fetters, and D. R. Hansen, *Rubber Chem. Technol.* **53**, 738 (1980).
47. M. Morton, *Anionic Polymerization: Principles and Practice*, Academic Press, Inc., New York, 1983.
48. M. Szwarc, M. Levy, and R. Milkovich, *J. Am. Chem. Soc.* **78**, 2656 (1956).
49. G. Holden, in L. M. Sperling, ed., *Recent Advances in Polymer Blends, Grafts and Blocks*, Plenum Press, New York, 1970, p. 276.
50. R. Moore, D. Russell, and W. C. Whechel, paper presented at the *ACS Rubber Division*, Toronto, Ontario, May 21–24, 1991.
51. H. R. Blum, in Ref. 36.
52. *Chem. Week*, 28 (Sept. 25, 1985).
53. R. J. Deisler, in Ref. 36.
54. G. Holden and K. H. Speer, *Autom. Polym. Des.* **8**(3), 15 (1988).
55. G. Holden and X-Y Sun, in Ref. 36.
56. D. J. St. Clair, *Rubber Chem. Technol.* **54**, 208 (1982).
57. G. Kraus, K. W. Rollman, and R. A. Gray, *J. Adhesion* **10**, 221 (1979).
58. Shell Chemical Co., Bulletin SC:198–83, Houston, Tex., 1984.
59. A. L. Bull and G. Holden, *J. Elastom. Plast.* **9**, 281 (1977).
60. D. R. Paul, in Ref. 1, Chapt. 12, Section 6.
61. M. J. Modic and L. A. Pottick, *Plast. Eng.* **47**(7), 37 (1991).
62. **U.S. Pat. 4,795,782** (Jan. 3, 1989), R. G. Lutz, R. Gelles, and W. P. Gergen.
63. E. J. van Beem and P. Brasser, *J. Inst. Petrol.* **59**, 91 (1973).
64. S. Piazza, A. Acrozzi, and C. Verga, *Rubber Chem. Technol.* **53**, 994 (1980).
65. J. L. Goodrich, "Asphalt and Polymer-Modified Asphalt Properties Related to the Performance of Asphalt Concrete Mixes," *Asphalt Paving*

Technologist Proceedings, AAPT, Vol. 57, 1988.

66. M. G. Bouldin, J. H. Collins, and A. Berker, *Rubber Chem. Technol.* **64**, 577 (1991).

67. E. I. du Pont de Nemours & Co., Inc., Hytrel Bulletin HYT-302R(1), 1981.

68. E. I. du Pont de Nemours & Co., Inc., Hytrel Bulletin IF-HYT-370.025, 1976; **U.S. Pat. 3,718,715** (Feb. 27, 1973), R. W. Crawford and W. J. Witsiepe (to Du Pont).

69. A. T. Chen, paper presented at the *3rd International Conference on Thermoplastic Elastomer Markets and Products*, Schotland Business Research, Dearborn, Mich., Mar. 28–30, 1990.

70. M. Szycher, V. L. Poirier, and D. Demsey, *Elastomerics* **115**(3), 11 (1983).

71. S. Abdou-Sabet, in Ref. 36.

Geoffrey Holden
Consultant



ELECTRICAL CONNECTORS

Electrical connectors are mechanical devices that connect wires, cables, printed circuit boards, and electronic components to each other and to related equipment. Connector designs include miniature units for microelectronic applications; specialized cable; rack and panel designs for incorporating combinations of a-c, d-c, and radio-frequency conducting contacts; and high current connectors for industrial application and for transmission and distribution of electrical power in overhead and underground networks. Further categorization of connectors can be made according to: application, whether connectors permanently join conductors and components or permit separation and rejoining; the means used to effect connection, whether by fusion (welding, soldering) or by pressure, the values of which can be small or great enough to severely deform metal; the distribution type, whether of power or of low (signal) levels of current; and the conductor size. The term electrical contact describes the junction between two or more current-carrying members that provide electrical continuity at their interfaces. Connector contacts ordinarily remain stationary in active circuits, eg, they are not mated or separated. Components having electrical contacts other than connectors include circuit breakers, switches, relays, and contactors that are designed to interrupt or to establish current flow in active circuits, and slip rings and brushes that transmit current from a stationary to a moving frame of reference.

Many connectors for single conductors have an insulating sleeve, and almost all connectors that join two or more conductors have a plastic body, or dielectric, which separates the contact elements (see INSULATION, ELECTRIC). Metal or plastic shells with mechanical aids, such as screws, levers, and other coupling devices to facilitate joining and separation of the contacts, also may surround a connector. The shell may have mounting features for securing the connector to a chassis, supports for the wires and cables, and polarizing keys for prevention of improper mating.

Connector Configurations

Electronic Connectors. The complexity and size of many electronic systems necessitate construction from relatively small building blocks which are then assembled with connectors. An electronic connector is a separable electrical connector used in telecommunications apparatus, computers, and in signal transmission and current transmission <5 A. Separable connectors are favored over permanent or hard-wired connections because the former facilitate the manufacture of electronic systems; also, connectors permit assemblies to be easily demounted and reconnected when inspection, replacement, or addition of new parts is called for.

The emergence of integrated circuit technology, which was characterized by the development of direct linkage of many circuit components or linkage via conductive paths on tiny ceramic substrates, was thought to threaten the growth of the electronic connector market (see INTEGRATED CIRCUITS). However, the explosive growth of electronics in most areas of manufacturing, business, and consumer products created new markets and resultant expansion of the connector field (see also ELECTRONIC MATERIALS; ELECTRONICS, COATINGS).

Electronic connectors may connect internally or externally. Internal connections may be between a component and a printed circuit board or wire (Fig. 1a); a printed circuit board and a wire or another printed circuit board which is in a chassis (Fig. 1b); and between chassis in the same cabinet (Fig. 1c). External connectors join separate pieces of equipment (Fig. 1d).

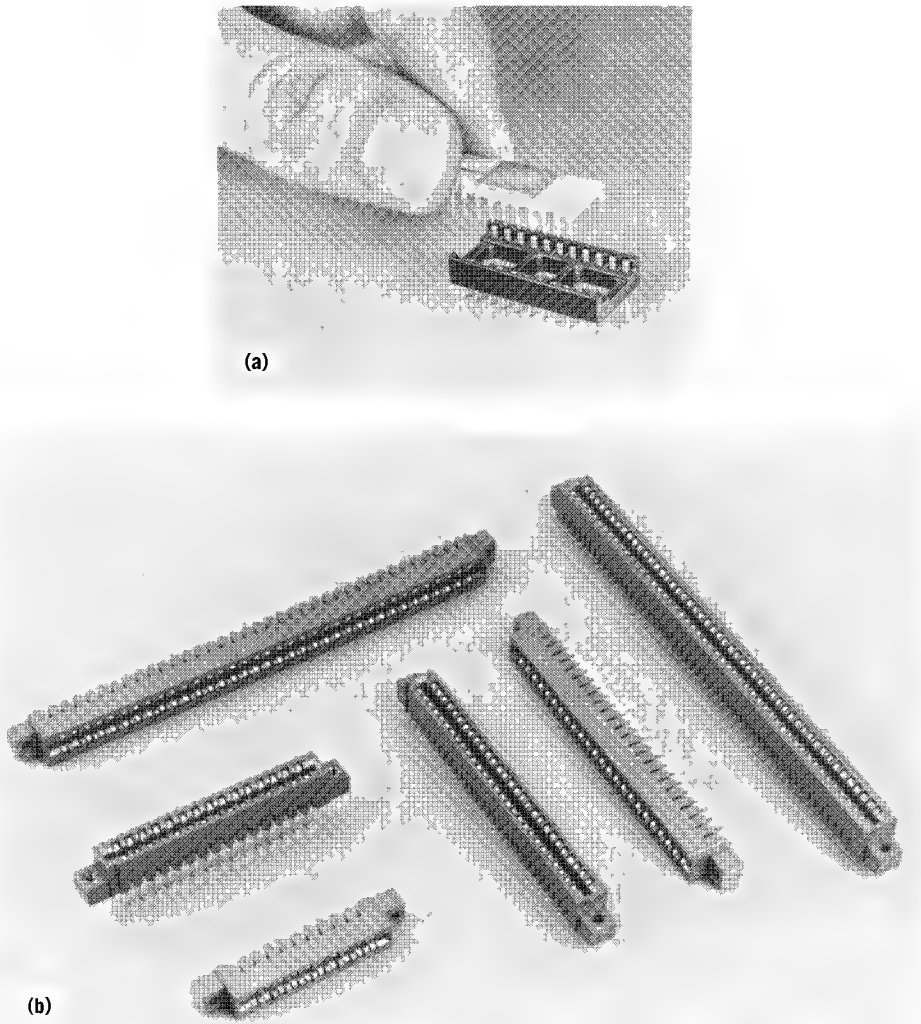


Fig. 1ab. Some types of electronic connectors. (a) Receptacle for dual-in-line package (DIP) semiconductor integrated circuit. (b) Connectors for printed circuit boards having edge contacts; two-piece connectors have male and female connector halves, one of which is attached to the printed circuit board,

usually by soldering. (c) Rectangular connector for chassis mounting. (d) Circular connector for cable. (a,d) Courtesy of Bunday Corp.; (b,c)

courtesy of AMP Inc.

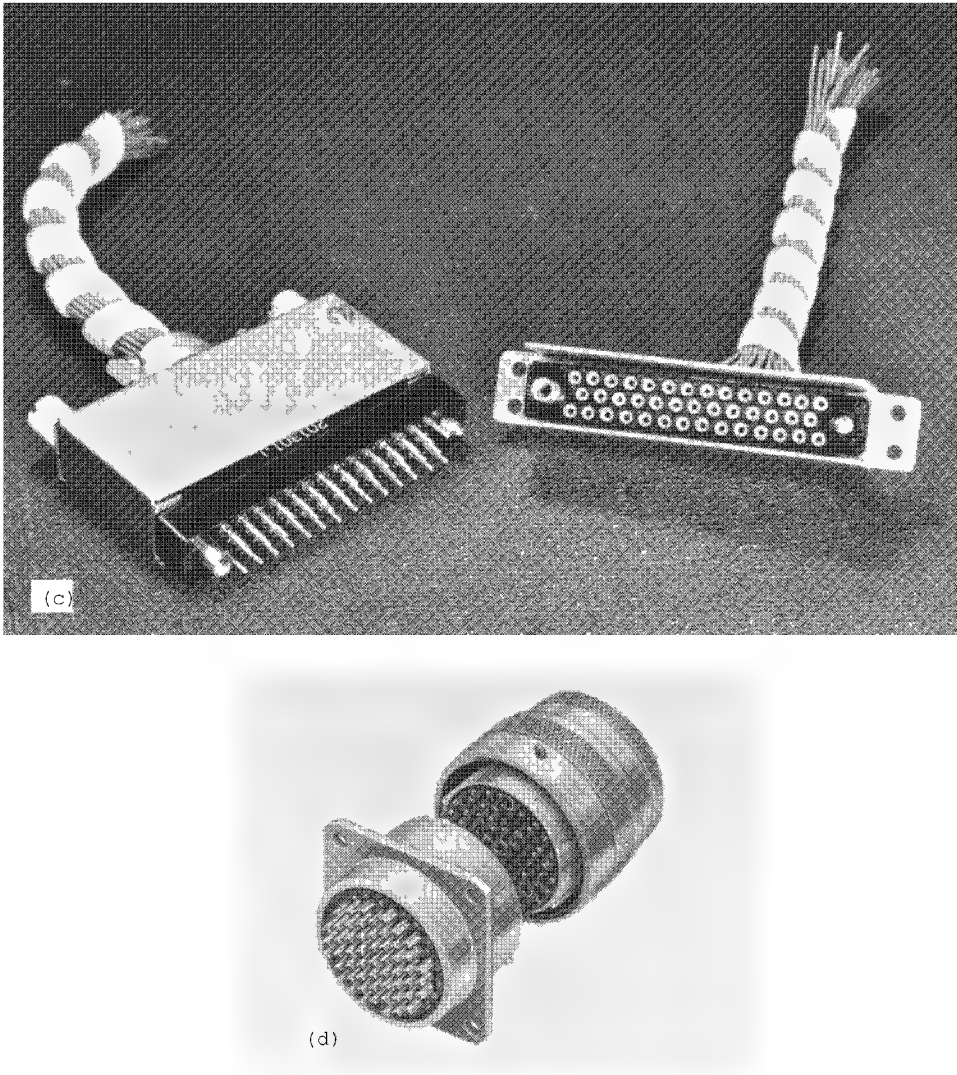


Fig. 1cd. Continued

A popular connection system consists of square metal pins, usually 0.064 cm (0.025 in.) in size, that are pressed into holes drilled in a printed circuit board. The holes are copper (qv) plated on the insides and interconnect conductors on the top and bottom faces of the board. Multilayer boards have interior circuits that may also be interconnected in this way. The pins have either a solid shank or a deformable (compliant) cross section where the pins join the board (Fig. 2). Separable connectors or solderless wraps (Fig. 3) engage the ends of the pins. One end of the pin can be the contact and spring of a separable connector.

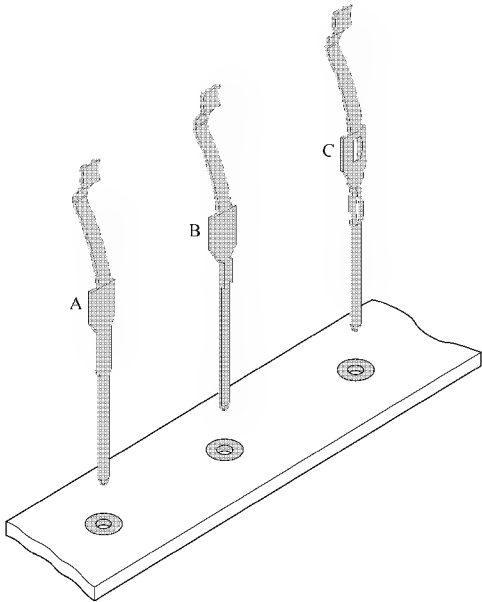


Fig. 2. Press-Fit pins of A, solid; B, crescent; or C, split-beam (compliant) design, are forced into through-holes of printed circuit boards where they interconnect conductors on opposite faces of the board.

Courtesy of AMP Inc.

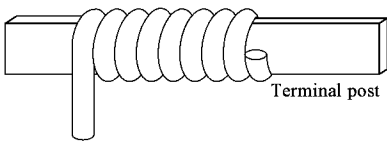


Fig. 3. Solderless wrap connection.

Surface mount refers to a method of securing connectors to the conductors of a printed circuit board by soldering appropriately shaped contacts to the board surface. Higher contact densities can be achieved and the need to drill holes in the board is avoided. Contact spacings may vary from about 0.5 cm for large current-carrying applications to 0.18 cm or less when miniaturization and high density is a requirement.

The contacts of an electronic connector have spring elements which press the mating surfaces together with a predetermined force, usually in the range of 0.25–5 N (0.056–1.12 lbf) for plated contacts; this range depends on the connector design and the materials from which the contact is made. Figure 4 illustrates typical spring designs. Mating of the connector is usually by sliding the surfaces together. Less frequently, the contacts are butted (Fig. 4h) after they have been positioned in zero insertion force (ZIF) connectors. Butting, which requires separation of contact springs by means of a cam and lever or equivalent, is preferred to sliding the halves together as a method of mating connectors having more than 100 contacts because it requires smaller friction forces. Butting not only reduces mechanical wear, especially when soft metal finishes such as tin plate (1) are used as the contact material, but facilitates connection to the fragile legs of some components (Fig. 1a).

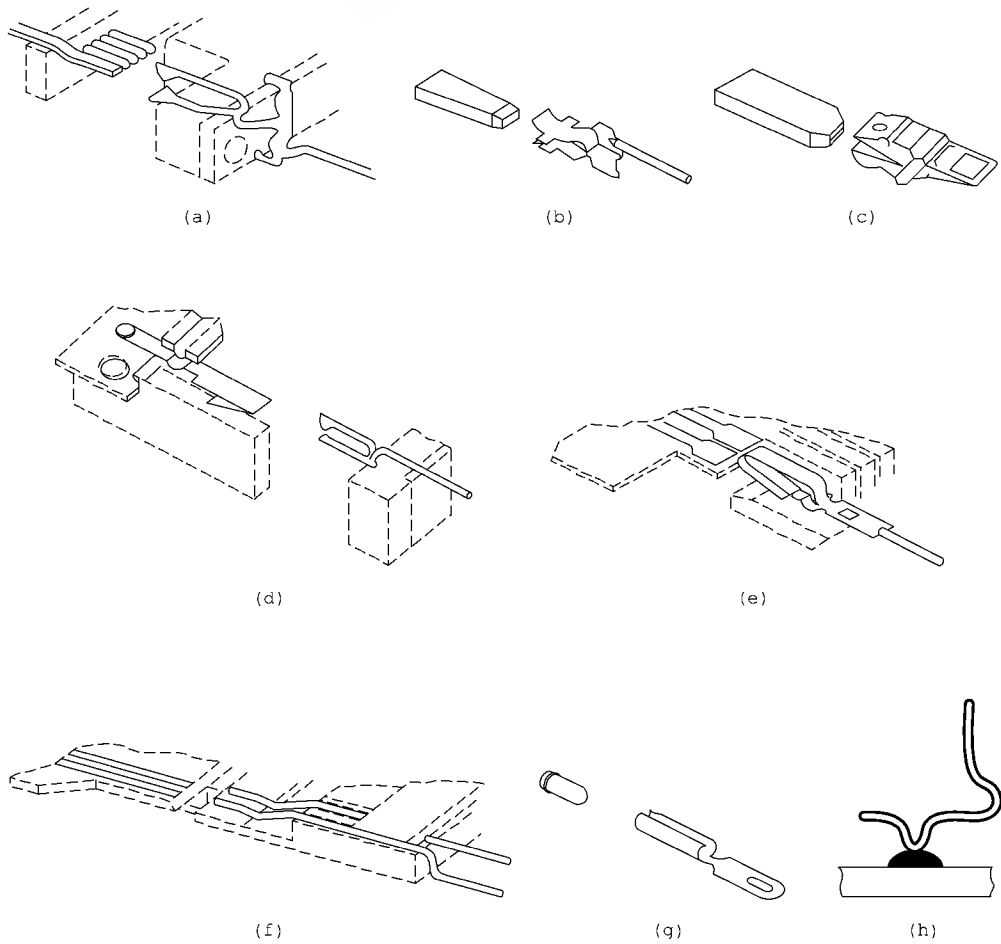


Fig. 4. Contact spring configurations for separable electronic connectors. Examples (a)–(g) are engaged by sliding the spring contact and its mating member together: (a)–(d) are blade-fork contacts; (e), folded cantilever contact for printed circuit boards; (f), straight cantilever contact for printed circuit boards; (g), pin-socket contact; and (h), butting contact. (a)–(g)

Courtesy of AT & T Bell Laboratories.

Another design which provides unusually low mating forces employs bundles of wires in both halves of the connector that intermesh, like two hair brushes, when the parts are connected (2).

Connector Shielding. Electromagnetic radiation, either human-made, ie, from tv, radios, radar, automotive ignitions, etc; or natural, ie, lightning, can interfere with the quality of signal transmission. Signals in conductors of electronic equipment including computers and communications gear may be seriously degraded. Methods used to control this effect include the use of coaxial cables and connectors in which shielding and grounding are used, twisted pair conductors, and filter connectors. An example (3) of this last is low pass filters for digital applications. These pass frequencies below a certain value and block any higher frequencies. One filter type consists of a special sleeve where the inside diameter is soldered to a center conductor such as a contact pin, whereas the outside diameter is soldered to a metal ground plane. Typical construction of the sleeve is given in Figure 5 and its use in a multicontact connector is shown in Figure 6.

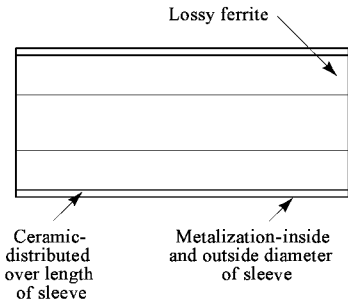


Fig. 5. Typical construction of a filter sleeve.

Courtesy of AMP Inc.

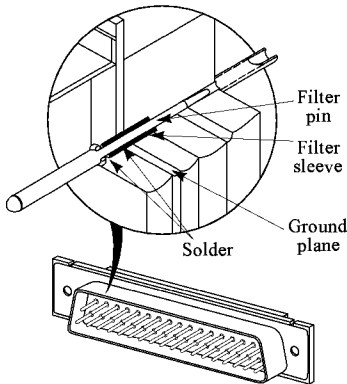


Fig. 6. Filter sleeve installed in a multicontact connector.

Courtesy of AMP Inc.

Another type of electronic connector joins coaxial conductors. These have a solid or stranded center-conductor surrounded by a dielectric. The dielectric is covered with a conductive shield made of metal braid or tape and with a layer of insulation. Coaxial cable connectors terminate the center-conductor and the shield. These are used primarily in radio frequency circuits. The shape, dimensions, and materials of an electronic connector shell or structure may have to be designed to shield the connection from electromagnetic and radio frequency interferences in many applications.

Joining to Electronic Connectors. The most widely used techniques for the termination of wires to separate contacts are the soldering (see *SOLDERS AND BRAZING ALLOYS*), welding (qv), crimping, solderless wrapping, and slotted-beam methods. Except for crimping and welding, it is usually possible to replace wires to a contact a limited number of times if repair or wiring changes are necessary.

Soldering materials are alloys that are composed primarily of tin and lead (qv), and have low melting temperatures relative to the conductor metals which are being soldered (see *LEAD ALLOYS*; *TIN AND TIN ALLOYS*). Welding requires sufficiently high temperatures for the fusion of metals.

Crimping is the compression of the back end of the separable contact, a tube, onto the wire conductor with a special tool that severely deforms both wire and barrel. This technique is suitable for joining both solid and stranded wires to connectors. Figure 7 illustrates typical crimp connections. Once the tool is removed the wire remains under the radial compression that is provided by the barrel. The force required to pull the copper wire from a crimped contact is approximately the same as its breaking strength. A soft metal, such as brass, is better for crimping than one having considerable springback, such as phosphor bronze (see *COPPER ALLOYS*).

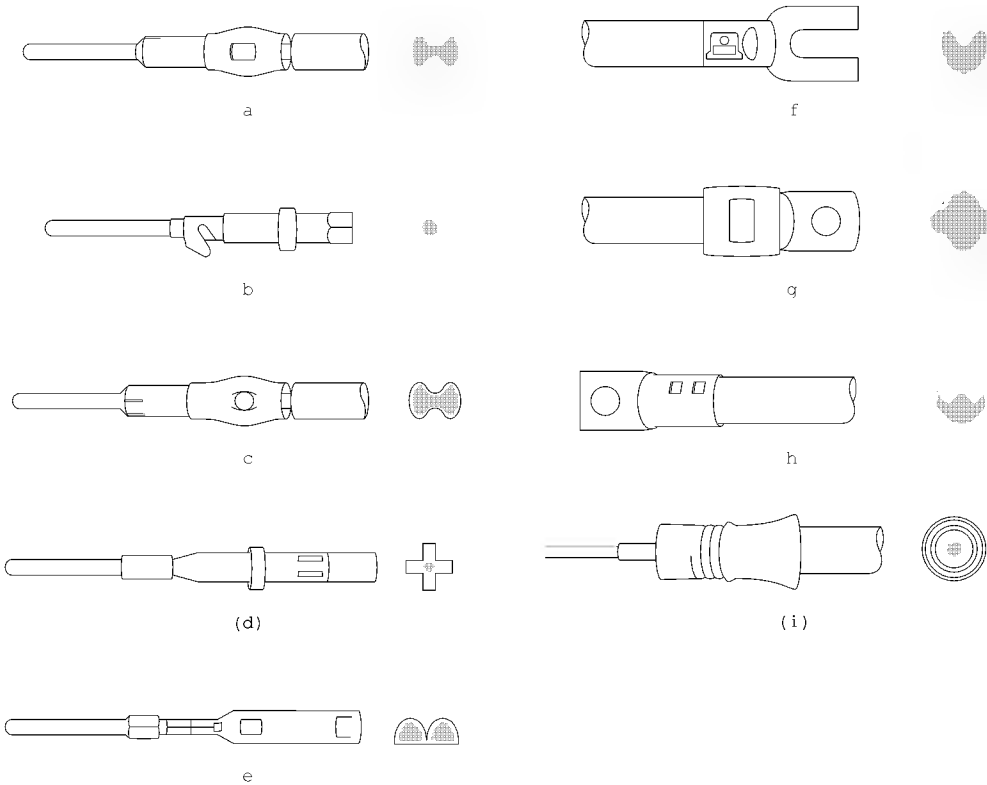


Fig. 7. Typical crimp contacts. Insets are cross sections of crimp indentations: (a) is longitudinal; (b), square; (c), double indent; (d), 4 indent; (e), B crimp; (f), nest indentor; (g) circumferential; (h), quad; and (i), interlocking.

Courtesy of Burndy Corp.

In the solderless wrap (Fig. 3) or wire-wrap connection, a wire conductor is coiled around the back end of the separable contact, which has a square or rectangular cross section (4). The corners of the solderless wrap post and the areas of the wire that are in contact with it are severely deformed. In a properly made wrap, the force required to slide the wire along the post exceeds the breaking strength of the wire. The method is suitable only for solid wire, and special tools are used to make this connection.

Slotted beam or U-contacts describe a versatile design for the termination of solid wire and require that the wire be pushed into a narrow slot between two moderately rigid tines, or beams, at the back end of the separable contact (Fig. 8). The edges of the beams displace the insulation, squeeze the wire, and keep it in compression for the life of the connection. This termination method was developed for terminating conductors in a gang using flexible flat cables with round conductors (5).

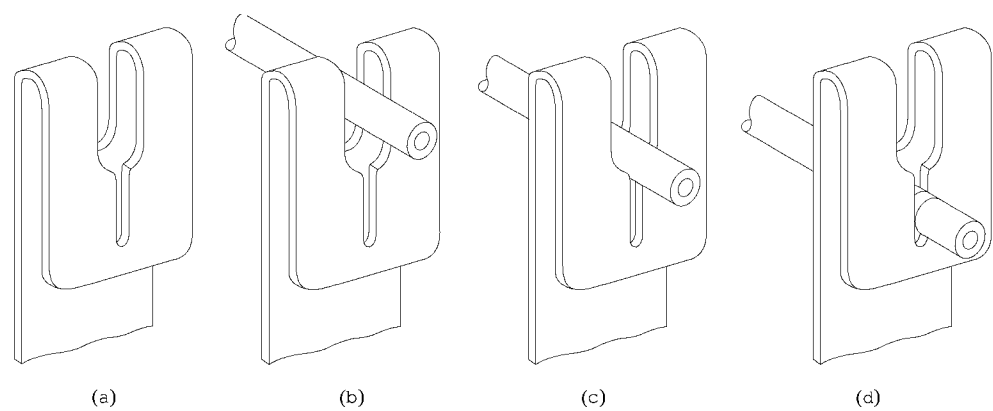


Fig. 8. Typical slotted beam connection where (a) is the terminal which resembles an inverted "U"; (b), the wire fits loosely into the upper portion of the slot; (c), the funnel shaped area displaces insulation; and (d), the conductor extrudes into the narrow bottom portion of the slot.

Courtesy of AMP Inc.

Methods used to secure a wire to the back end of a separable contact include the taper pin and the solderless clip. The former is a cylindrical tapered body having a hollow end into which a wire is crimped; the front of the taper pin is forced into the back end of the connector contact which has conforming shape. The solderless clip has a spring which traps the solid or stranded wire against a post at the back end of the separable contact; the clip encircles both the wire and the post (6).

Splicing Connectors. Splicing connectors are used to permanently join wire to wire. Some are simple sleeve barrels that are crimped to bare wire; others are preinsulated where the crimp is made by compressing the sleeve and its positioned insulation onto wires which may or may not have insulation. Two types of splicing connectors, which require insulation displacement and are used in the telecommunications industry, are illustrated in Figures 9 and 10. The connector in Figure 9 has a spring-tempered phosphor bronze liner that pierces the insulation and contacts the wire and an outer shell of soft brass (7). This connector is suitable for paper-insulated solid conductors. The slotted-beam connection principle is exemplified in the connector shown in Figure 10. This connector gang-terminated 25 pairs of wires using either paper or plastic insulation (8). The conductor can be aluminum or copper and its diameter ranges from 0.40–0.81 mm (20–26 American Wire Gage). A grease sealant is used to provide moisture resistance.

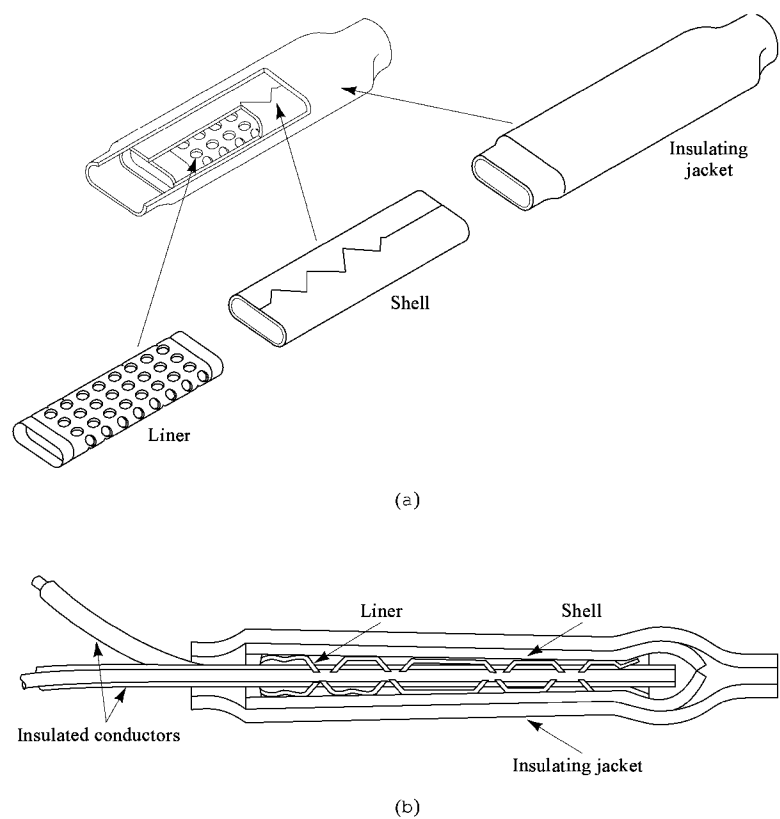


Fig. 9. Splicing connector for two wires (insulation piercing). (a) Exploded view; (b) cross section after pressing on wires.

Courtesy of AT & T Bell Laboratories.

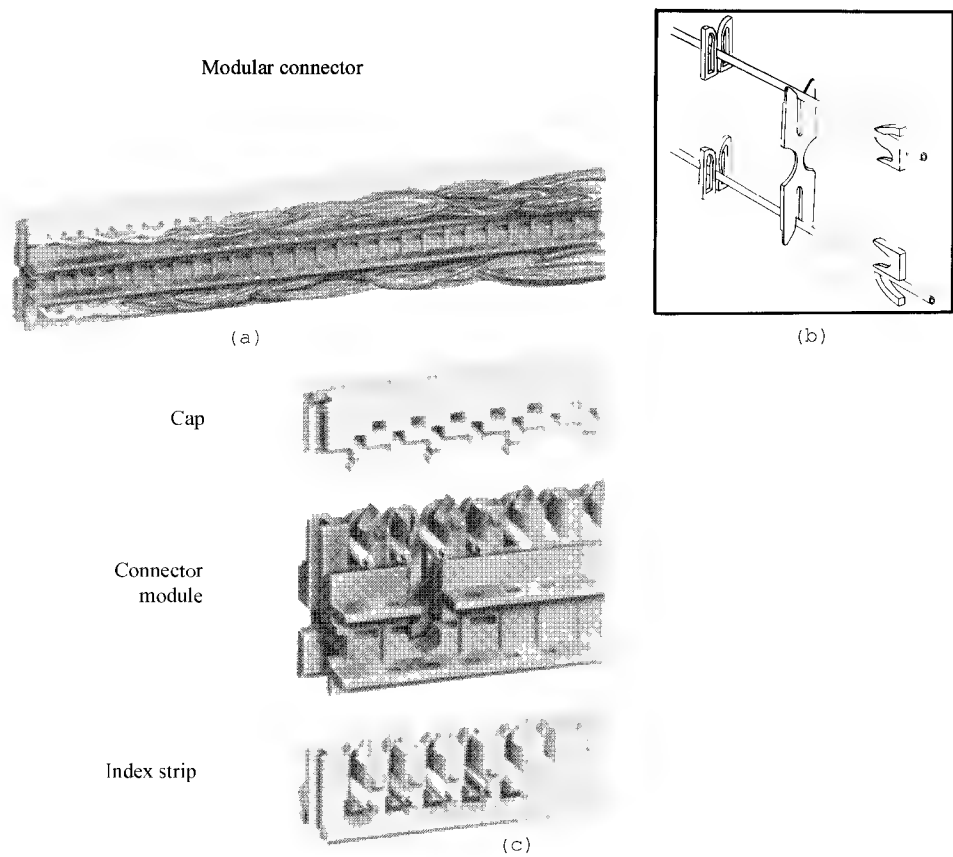


Fig. 10. (a) Splicing connector for terminating 25 pairs of telephone wire; (b) schematic of the slotted-beam insulation piercing contact; and (c), components of the modular connector.

Courtesy of AT & T Bell Laboratories.

Terminals. Terminals are connectors having individual wires that are designed to be screwed down at separable ends, and to which conductors are permanently joined at the back end, usually by crimping. Figure 11a,b,c illustrates common terminal configurations. Either ring or open-tongue configurations provide a terminal for the screw connection. Another popular terminal, called a quick disconnect (Fig. 11d), has a spring receptacle into which a contact tab with a rectangular cross section is inserted. This latter design is widely used in the appliance industry.

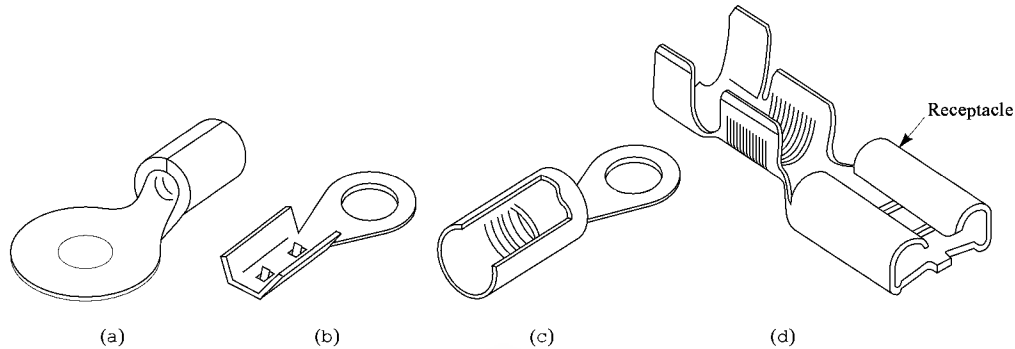


Fig. 11. Crimp terminal configurations: (a) straight barrel, 90° tongue, where wire without insulation is crimped in the barrel. (b) Open barrel having insulation-piercing lances. (c) Nylon or poly(vinyl chloride) preinsulated terminal accommodating and supporting wire insulation. Wire without end insulation is inserted in the terminal and is crimped. The terminal sleeve is not broken but conforms to the shape of the crimp indent. (d) Quick disconnect terminal having a crimp barrel and provision for insulation support.

Courtesy of AMP Inc.

Utility and Industrial Connectors. Connectors used in power distribution systems are nearly always of the permanent type and are usually made for single conductors. Sleeve barrels are used to splice cable by crimping (Fig. 12a). Insulating covers may be applied to the connector after the joint is made between the connector and the insulated cable. Clamp-type connectors are used with one or more clamping bolts and may have additional resilience when used with dished washers made of spring steel (Belleville washers). Figure 12b illustrates a typical clamp connector in a terminal configuration which is used for insulated conductors in building construction applications. Most crimp power connectors for aluminum and copper cable are made of aluminum. Those that are fabricated from copper alloys are for copper cable only. Clamp connectors are made of copper and aluminum alloys.

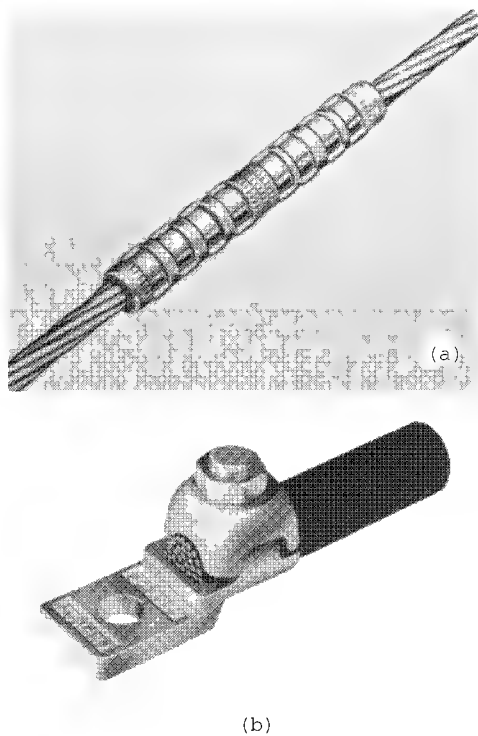


Fig. 12. Typical connectors for power distribution cable. (a) Aluminum sleeve barrel for uninsulated cable. (b) Clamp-type terminal connector for insulated cable used in a building construction.

Courtesy of Burndy Corp.

Methods of Application. Attachment of separable contacts to conductors may be achieved using automated machinery or specialized hand tools. The automated machinery method is popular in large-volume original equipment markets for products such as strip terminals, rack and panel connectors, and printed circuit-board connectors. This machinery applies contact components to conductors, but the insertion of the conductors and the applied contacts into connector housings is usually performed manually. The cases in which the desired connection is not amenable to automation require the use of hand-tooled application. Products included in this category are heavy-duty terminals and splices, connectors to which conductors are terminated by soldering, and small wire terminals. Hand-tool application is typically associated with utility, construction, maintenance and repair, and small-volume original equipment markets. The largest connectors, which are employed by utilities for construction and servicing of power distribution lines, require installation tools powered by hydraulic means, compressed air, or small explosive charges.

Contact Principles

The design of electrical contacts and the selection of contact materials are based on a body of interrelated physical, metallurgical, and chemical principles. In the early 1920s, Ragnar Holm pioneered studies on the origin of resistance to current flow at the surfaces of touching metallic bodies. From those investigations contact science emerged as a specialized field of physics and materials technology.

Nature of Mechanical Contact. The surfaces of solids are irregular on a microscopic scale. Even nominally plane, smooth surfaces have a large-scale waviness on which is superimposed a roughness having peak-to-valley distances of several micrometers. When two metallic bodies are placed in contact at a light load, the bodies touch at only a few small spots, or asperities. As the load is increased, more and more asperities come into contact and the surfaces move together. The true area of contact depends, therefore, on normal load and on the hardness of the metal. The real area of contact is only a fraction of the apparent area in most cases, except at very high loads. For example, the ratio of real to apparent contact areas of finely lapped steel flats having an apparent area of one cm² is ca 10⁻⁴ at a force of 10 N (2.25 lbf).

Nature of Electrical Contact. If metallic surfaces are covered by a nonconducting layer, such as an oxide or a sulfide tarnish film, the area of metallic contact is zero provided that the film is unbroken. Significant current does not flow between such surfaces except when the film is less than 2–3 nm in thickness. At or less than these thicknesses, electron tunneling (tunnel conduction), which is voltage independent, occurs by a wave-mechanical effect analogous to the transmission of light through metal foil of thickness comparable to the wave length (9). If the nonconductive layer on a surface is discontinuous or is punctured, the mechanical load is borne by both film and metal. Current then flows through the metallic spots, called a spots (10). The lines of electric flow converge at these spots, as illustrated schematically in Figure 13. Constriction resistance, the increase of resistance beyond that of a continuous solid, ie, not having an interface (11), originates at this convergence.

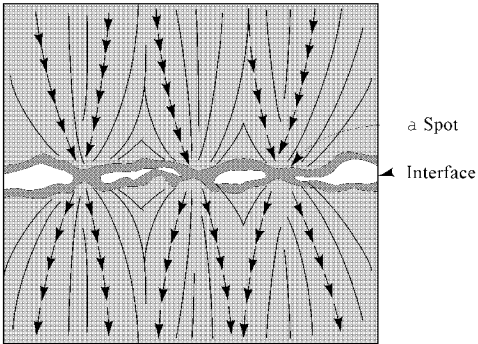


Fig. 13. Schematic of a microscopic view of contact interface where constriction resistance originates in the constriction of current flow through the touching metallic junctions (a spots) of the mating surfaces. The arrows and lines indicate the flow of current.

In the simplest case, for a single circular contact spot between identical metals having a uniform film, contact resistance R has the relationship:

$$R = R_c + R_f$$

$$R = \frac{\rho}{2a} + \frac{\sigma}{\pi a^2}$$

where R_c = constriction resistance, R_f = film resistance, a = radius of the a spot, ρ = bulk resistivity of the contact metal, and σ = film resistance (ohms per unit area). The radius of the a spot can be calculated from the hardness of the metal and the applied load.

Because a length of metal associated with the connector contact is ordinarily in the path between the contact end to which a wire is terminated and the contact interface, its resistance (bulk resistance) must be added to contact resistance when considering the connector as a circuit element. This overall resistance is sometimes erroneously called contact resistance.

Resistance Heating of Contacts. The contact material, contact area, and heat dissipating ability, as well as the heat dissipating ability of the structure to which the material is attached, limit the amount of current that a contact can transport. Excessive current heats and softens the metal contact. This softening results in an increase in the surface area of the contact and a corresponding reduction in contact resistance.

At higher currents, the mating junction melts. Both softening and melting occur at characteristic voltages. Typical values are shown in Table 1.

Table 1. Softening and Melting Voltages^a

Metal	Softening		Melting	
	°C	V	°C	V
Sn	100	0.07	232	0.13
Au	100	0.08	1063	0.43
Ag	150–200	0.09	968	0.37
Al	150	0.1	660	0.3
Cu	190	0.12	1083	0.43
Ni	520	0.22	1453	0.53
W	1000	0.6	3380	0.1

^a Ref. 12.

Voltage Breakdown of Films. If a significant voltage can be passed by a circuit across a film-covered contact, the film, depending on thickness and composition, may break down electrically. This action, called the coherer effect or fritting (13), results in the formation of minute metallic conductive paths through the film. Puncturing is of the order of 0.1 V /nm of film. The potential drop across the film after puncturing is the melting point voltage.

Overview. Metallic contact between surfaces of separable connectors usually is obtained either by using noble metals, which are essentially film-free, or by designing the contact so that any films that are present are broken before or as the surfaces are brought together. For example, noble metal coatings on base metal substrates having high conductivity, such as copper alloys, are used for small multicontact connectors which can provide small mechanical loads normal to the surface. Alternatively, soft metal contacts of, for example, tin and tin–lead alloy platings can be used because films on their surfaces are easily disrupted on closure. In the latter case, mechanical wear from sliding connector designs may severely limit the numbers of insertions and withdrawals that are possible. In the case of power connectors, severe deformation of the aluminum or copper surface is obtained by crimping or clamping facilitating disruption of the oxides on both the connector and conductor surfaces. Tin plate is used widely on aluminum connectors, and wire brushing of the cable (especially if it is made of aluminum) and joint aids increase the areas of metallic contact and thereby lower contact resistance. A large part of practical connector engineering is devoted to designing metallic contacts that need minimal maintenance, especially for those intended to serve in chemically aggressive atmospheres and in high temperature applications.

The contact resistance of any electrical connector in a circuit must be stable and generally low for proper functioning of that circuit. Low voltage circuits, which are common in modern electronic systems, have open-circuit voltages of not more than a few volts. These are insufficient levels for the fritting of films that grow on base metals from environmental exposure. Metallic bridges established by fritting are fragile, and the voltage drop across fritted surfaces may permit heating and consequent degradation of the contact. It is, therefore, generally undesirable to rely on fritting as a method that may establish current flow in a connector. Low contact resistance should be achieved using noble metal contacts or base metals concurrently with methods that mechanically perforate any insulating films that are present.

Typical separable electronic connectors have contact resistances that range from several milliohms to tenths of ohms. Utility–industrial connectors, which are established by very high pressure connection and are exemplified by a bolted joint, have contact resistances of microohms. Crimping, solderless wrapping, and other methods used to establish connection of wires and other conductors to electronic connectors have associated contact resistances of ca 10–100 μΩ. The bulk resistance of the conducting elements, which include connector parts such as the spring, barrel, and pin, is usually several times greater than the contact resistances of the separable contact interface and the terminal ends. The overall resistance of electronic connectors generally ranges from 3–30 mΩ.

Materials and Processing

Contact Substrates. The substrate must be able to be terminated readily as well as be a good electrical conductor. In electronic connectors the substrate may serve as a spring element. The most widely used spring materials for connectors are the copper alloys: 98.1 wt % Cu; 1.9 wt % Be (Unified Numbering System (UNS) designation C17200); 94.8 wt % Cu, 5.0 wt Sn, 0.19 wt % P (phosphor bronze, C51000); and 88.2 wt % Cu, 9.5 wt % Ni, 2.3 wt % Sn (C72500) (see COPPER ALLOYS). Sometimes springs made of metals that have poor conductivity, such as stainless steel, are used as inserts in connector barrels of brass or similar metals which are inexpensive and can be terminated easily.

The contact ends of printed circuit boards are copper. Alloys of nickel and iron are used as substrates in hermetic connectors in which glass (qv) is the dielectric material. Terminals are fabricated from brass or copper; from nickel, for high temperature applications; from aluminum, when aluminum conductors are used; and from steel when high strength is required. Because steel has poor corrosion resistance, it is always plated using a protective metal, such as tin (see TIN AND TIN ALLOYS). Other substrates can be unplated when high contact normal forces, usually more than 5 N, are available to mechanically disrupt insulating oxide films on the surfaces and thereby assure metallic contact (see CORROSION AND CORROSION CONTROL).

Contact Finishes. Base metal substrates quickly develop oxides, tarnish, or corrosion films, in humid, polluted atmospheres. Because these films may prevent adequate metal-to-metal contact when the connector or connector–conductor surfaces are mated, coatings of other metals commonly are used to obtain corrosion resistance, to provide conductivity, or to facilitate termination to conductors by soldering, wire wrapping, or by other means. Application of finishes is achieved by electroplating (qv), cladding, and by hot-dipping when low melting metals such as tin are used (see METAL SURFACE TREATMENTS). Selective application of a contact finish to portions of the substrate can be accomplished using any of these coating techniques. The principal noble contact finishes are gold, palladium, rhodium, and alloys having a high gold or palladium content. Palladium-based finishes usually have a thin (0.05–0.1 μm) top coating of gold. The non-noble finishes are tin, silver, and nickel. Alloys of these metals, such as 50 wt % Sn, 50 wt % Pb, also are widely used.

Noble Metal Electroplated. Because it is chemically stable, gold is the noble metal most extensively used in the contacts in electronic connectors; metric ton quantities are consumed annually for this purpose. The high cost of noble metals requires that these materials be sparingly, however effectively, used. The chief requirement of a plated finish is sufficient thickness. Gold coatings in thicknesses of 0.1–2 μm are used; the greater thicknesses are required for critical applications and for wear resistance when large numbers of engagements of the connector are specified. Rhodium ordinarily is plated from only 0.5–1 μm. Thicker coatings have a tendency to crack spontaneously. The thickness of palladium and palladium–nickel (commonly the 80 wt % Pd, 20 wt % Ni alloy) electrodeposits parallels that of gold.

Porosity ranks next to thickness in importance, especially when the finishes must serve in polluted and or humid environments which promote tarnish and corrosion. Pores, openings in the surface that extend to the underplate or substrate, can be intrinsic in the coating (14), or can be produced by mechanical wear or by forming operations involved in manufacturing. In some environments the substrate can tarnish or corrode at pore sites and can produce localized areas of insulating films which cause contact resistance to increase. Porosity is less important for connectors that operate indoors at moderate to low relative humidities and in the absence of corrosive pollutants (15).

Hard deposits enhance a contact's resistance to mechanical wear. Palladium, palladium–nickel alloy, and gold electrodeposits containing 0.1–0.5 wt % cobalt or nickel are more wear resistant than pure gold finishes (16). Deposits must be solderable if the associated terminations are to be soldered. Gold deposits thicker than 2.5 μm form brittle gold–tin intermetallics (17) during soldering and, therefore, give weak joints. Contaminated surfaces are also difficult to solder.

Underplatings normally are used for noble metal electrodeposits. Zinc from brass rapidly diffuses through gold plate and forms insulating films on the plated surface. Its rate of diffusion is reduced by copper or nickel underplatings (18). Diffusion of copper through gold plate, especially at temperatures above 100°C, can be troublesome, and nickel underplate is used to retard the diffusion rate (18). Copper and nickel underplatings are usually used to lower porosity in a noble metal deposit (14). The effective hardness of a multilayer coating is increased by hard underplates. The wear resistance of gold deposits can be improved by nickel underplating (19). Pure electroplated nickel is ductile and harder than most substrate metals, and the thickness of the nickel underplate should be at least 1.25 μm in order to significantly reduce wear. The corrosion tendency of contacts that have porous noble metal finishes is lowered by nonreactive underplatings, such as 65 wt % Sn, 35 wt % Ni (20).

Clad Inlay Noble Contact Materials. Insertion of a strip of metal into a groove in a base metal substrate which is then metallurgically bonded to the substrate by rolling at high pressures (Fig. 14) is referred to as cladding (21). The metals must be ductile, and some copper alloys, such as C72500 and C15000, which are used in connector springs, are especially suitable for this purpose. Pure gold, pure palladium, and alloys of gold and silver or nickel, such as 75 wt % Au, 25 wt % Ag; 96 wt % Au, 4 wt % Ni; and 60 wt % Pd, 40 wt % Ag having a thin gold cap which has been partially diffused into the palladium alloy by heating (22), are widely used clad inlays for electronic connector contacts. Because intermediate heat treatments as well as several passes through the rolls are usually necessary in order to achieve the desired thickness reduction, a nickel interliner is ordinarily used as a diffusion barrier. The same thickness and porosity requirements for electrodeposits apply to clad noble metal finishes.

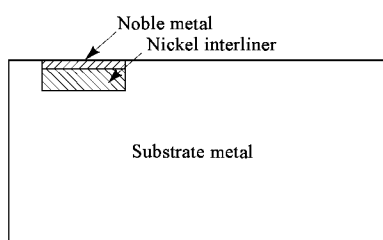


Fig. 14. Schematic of a clad inlay contact material.

Cladding may be less expensive than selective electrodeposition when coatings greater than 1 μm of a noble metal are required, but may be more expensive than electrodeposition for thinner coatings. Selective techniques are most easily used for sheet metal substrates that are to be machine stamped and formed into contacts. Clad noble metals are considerably more ductile (and less hard) than comparable electrodeposits and, therefore, are better suited to forming operations. Contacts that are made into separate parts from rod by screw machining are usually coated on all exposed surfaces by barrel electroplating.

Noble Metal Weldments. Noble metal contact buttons, 25–75 μm thick, are made by resistance welding a rod of the material to the substrate, which usually is a contact spring, and by cutting the rod and forming the button to the desired shape. Pure gold and gold–silver alloys are the most commonly used metals.

Base Metal Finishes. The low cost of base metal finishes obviates selective coating. Electrodeposition is used for 0.5–5 μm thick coatings of tin and tin–lead alloy, usually about 50 wt % Sn, 50 wt % Pb, on electronic connector contacts, on contacts at the edges of printed circuit boards, and on terminals. Sheet copper alloys that have been coated with tin–lead alloy are widely used for contacts that are stamped and then formed into the desired shapes, such as pins having a closed end and sockets. Aluminum connectors that have utility–industrial applications are more thickly coated, and hot-dipping in molten tin is common.

Whiskers are filamentary growths a few micrometers in diameter that occur spontaneously from tin and other low melting pure metals in response to lattice strains. Whiskers can attain lengths of several millimeters and may short-circuit adjacent contacts. The presence of whiskers is unimportant if currents are in the ampere range, because bridging is instantly cleared by melting. Whisker formation can be retarded by combining small amounts of alloying elements, such as lead or copper, with tin coatings or by fusing the coatings after plating (23).

Conductive Elastomers. Conductive elastomers (qv) (24), rubbers that are made conductive by molding metal or carbon powders in them, have characteristics of both a contact material and a spring. Silicone rubbers, neoprene, polyurethane, and other elastomers have been used; however, silicones are the most popular because these have a low compressive set and operate over a wide temperature range, from ca 65 to 200°C (see also ANTISTATIC AGENTS; ELECTRICALLY CONDUCTIVE POLYMERS). Particle loadings are high, eg, 70 wt % or more for metals, because there must be particle contact through the body for the elastomer to be conductive. Silver is used in those contacts that must be highly conductive and other metals are used in systems where higher resistance can be tolerated.

However, conductive elastomers have only ca $<10^{-3}$ of the conductivity of solid metals. Also, the contact resistance of elastomers changes with time when they are compressed. Therefore, elastomers are not used where significant currents must be carried or when low or stable resistance is required. Typical applications, which require a high density of contacts and easy disassembly for servicing, include connection between liquid crystal display panels (see LIQUID CRYSTALS) and between printed circuit boards in watches. Another type of elastomeric contact has a nonconducting silicone rubber core around which is wrapped metalized contacts that are separated from each other by insulating areas (25). A newer material has closely spaced strings of small spherical metal particles in contact, or fine solid wires, which are oriented in the elastomer so that electrical conduction occurs only in the Z direction (26).

Contact Lubricants. Debilitating wear of separable connector contacts, which may occur if the metallic coating is thin or if forces normal to the contact surfaces are high, can be minimized by coating the contacts with thin films (qv) or organic lubricants (16,27) (see LUBRICATION AND LUBRICANTS). Viscous mineral oils, poly(phenyl ethers), per(fluoroalkyl) polyethers, soft microcrystalline waxes (qv), and petrolatum have been used on electronic connector contacts. Although these lubricants are insulators, the few asperities of the metal pieces that make contact through the film provide low contact resistance, which is indistinguishable from that of unlubricated contacts. Some lubricants have corrosion inhibiting or antitamishing properties for base metals and for porous noble metal finishes (16).

Joint Aids. Good metallic contact with aluminum connectors or aluminum conductors, including those with severely deformed surfaces, is difficult if aluminum oxide is present. It is believed that contact occurs by extrusion of the substrate metal during deformation through minute cracks produced in the oxide; and that some degree of geometric matching of cracks is necessary (28) when both of the surfaces are aluminum. It has been found that finely divided zinc incorporated in grease can be coated on the interior surfaces of an aluminum connector to provide lower initial contact resistance and better long-term resistance stability (29). The grease is also able to retard ingress of air and water which can seriously degrade reliability of aluminum connectors, especially in exposed out-of-doors service in power distribution networks. Other soft metal coatings are effective, such as indium, which has been proposed (30) as a plating on copper alloy connectors intended for permanent joints to aluminum communications cables having slotted-beam connectors.

Insulators. Molded plastics serve as insulators for multicontact connectors and glass is used in hermetic connectors intended for bulkhead

mounting. Environmentally sealed circular connectors combine both elastomeric technology with plastics by bonding the seals and grommets to the plastic insert. A wide variety of plastics are employed in electronic connector bodies depending on the size, strength requirements, complexity of the design, and service environment. Plastic materials are often reinforced using about 30 to 40° glass fibers. Thermoplastics are favored by a wide margin over thermosets for electronic connectors because of manufacturing economies and generally superior performance.

Automated soldering operations can subject the molding to considerable heating, and adequate heat deflection characteristics are an important property of the plastics that are used. Flame retardants (qv) also are often incorporated as additives. When service is to be in a humid environment, it is important that plastics having low moisture absorbance be used. Molding precision and dimensional stability, which requires low linear coefficients of thermal expansion and high modulus values, are key parameters in high density fine-pitch interconnect devices.

Nylon and poly(vinyl chloride) sleeveings are used for preinsulated terminals. Ceramics (qv) are employed in some high voltage power connectors. Hard rubber shells insulate connectors that serve underground power distribution cables.

Some of the common types of plastics that are used are thermoplastics, such as poly(phenylene sulfide) (PPS) (see POLYMERS CONTAINING SULFUR), nylons, liquid crystal polymer (LCP), the polyesters (qv) such as polyesters that are 30° glass-fiber reinforced, and poly(ethylene terephthalate) (PET), and polyetherimide (PEI); and thermosets such as diallyl phthalate and phenolic resins (qv). Because of the wide variety of manufacturing processes and usage requirements, these materials are available in several variations which have a range of physical properties.

Reliability and Testing

Connectors must be as reliable as any component in the circuits that they serve. Reliability requirements of each connection are particularly stringent in complex apparatus or where signal and power must be carried long distances along a series of connector contacts. The consequences of failure include not only loss of service, but the great expense of locating the failure, exposing the connection, and making a replacement.

Considerable effort has been directed to determining the causes of connection failures and to learning how to minimize the likelihood of occurrence. Acceptable failure rates range from <1 in 10⁹ operating hours for contacts in air-frame (31) electrical systems and in some telecommunications equipment, to 100–1000 in 10⁹ operating hours in instruments, to even larger rates for contacts in many consumer products. A failure is defined as exceedance of contact resistance, which can be as little as twice the initial contact resistance, that causes circuit malfunction. The required lifetimes of connectors may be >20 yr, although most required application times are shorter (see MATERIALS RELIABILITY).

Mechanisms of Failure. The causes of connector contact failure can be of a thermal, chemical, or mechanical nature, in addition to misapplication and physical abuse.

Thermal. If a connector is in an environment hotter than that for which it was designed, or if it undergoes significant resistive heating resulting from excessive current flow, the following can occur and eventually may cause failure: (1) stress relaxation of the contact spring, or creep of the deformed metal, as with crimp barrels, which are responsible for maintaining pressure contact; (2) acceleration of chemical reactions that cause failure; for example, rapid thickening of the oxide on base contact surfaces may occur even when they are mated if the contact is not gas tight, eg, if the contact is unable to resist encroachment of the atmosphere at the mating site; (3) accelerated diffusion of substrate metals or of base hardener and impurity metals through the surface of noble metal platings and the formation of insulating films; and (4) volatilization of connector contact lubricants.

Alternating temperature cycles may be more damaging than a constant elevated temperature because differences in the amount of expansion and contraction of the members supporting the contacts can cause breakage of the metallic junctions (a spots). When there is an insulating film on the surface between the a spots as with base metals, metallic contact cannot be maintained owing to the movements. Heat cycling forms the basis of much connector testing, eg, 500 cycles of heating by current flow to 100°C above ambient followed by cooling to room temperature for aluminum conductor overhead power distribution connectors (29). A connector is acceptable if this stress does not cause significant change in overall resistance.

It is not practical to determine the contact resistance in power connectors. The resistance of the connection of a specified length of conductor on each side of the connector is measured and is called the overall resistance or the connection resistance. One industry specification (32) defines the included lengths and requires the stability of the connection resistance to be within ±5% of its average value throughout the heat-cycle test.

Chemical. Contacts made of base metals and contacts coated with noble metals and having pores may be subject to chemical reactions with air pollutants which cause formation of insulating films. Above a critical humidity, usually 70° , galvanic corrosion occurs at pore sites if the substrate is corrodible. Tarnish films, which may form if copper or silver reacts with elemental sulfur or with H₂S, may spread (33) from the pore sites to the remaining surface of the noble metal. Soils such as hygroscopic dust contamination and fingerprints lower critical humidity. SO₂ and certain reactive chlorine-containing pollutants, such as HCl, are aggressive at parts per billion levels.

Another humidity related degradation is silver migration from silver conductors (34): the transport of silver from one conductor to another through or on the surface of the insulator under the influence of a d-c electrical field. The positively charged conductor loses silver by oxidation in the form of positive ions, which then move through moisture paths to the cathode where they are reduced to metallic silver. The silver may grow in dendritic form and can eventually short-circuit the conductors. Current leakage between adjacent conductors through absorbed films of moisture may occur, especially if ionic contaminants are present.

Chemical degradation can be avoided by using closed structures, which may have protective covers or which may be fully hermetic, as well as by using unreactive metals. Barrier coatings are effective in some cases, and polyethylene–polybutene grease (8) is used in some splicing connectors for telecommunications cable.

Many techniques have been developed for the accelerated testing of connector contacts. These include elevated relative humidity exposure, cycling temperature–humidity procedures, and aging in chambers containing gaseous pollutants. The International Electrotechnical Commission has published a standard method for testing connectors which involves exposing the connectors for 4, 10, or 21 days in a chamber to a flowing stream of air containing 25 ppm of SO₂ at 75° rh and 25°C (15). Tests such as these, which are based on single corrosive gases, are being replaced by procedures that involve mixtures of three or four gaseous pollutants, eg, Cl₂, H₂S, NO₂, and SO₂, each at fractional ppm levels (35). Because of the complexity of environmental effects, age acceleration factors with respect to real conditions are generally unknown for tests in current use. Nevertheless, these serve as a design aid for estimating the relative merits of candidate connector contact materials and connector structures and for qualifying products.

Mechanical. Premature wearout or loss of contact metal during engagement and separation can result in loss of tolerances, reduced spring forces, formation of loose metallic wear debris, which may short-circuit contacts, and development of porosity in noble metal contacts. Underplatings, contact lubricants, and hard materials reduce mechanical wear.

Fretting corrosion (36,37) can lead to high contact resistance of base metal contacts, such as tin plate in electronic connectors. Small cyclical displacements of the connector halves occur because of external vibration or differential thermal expansion and contraction of the mating contacts. The wear debris that is formed remains in the contact zone. The accumulation of oxide debris in the contact region leads to increased contact resistance. Solutions to this problem are structures that do not permit movement of contact surfaces with respect to one another, the use of gold as a contact finish, and the application of thick coatings of contact lubricants and greases, which reduce the rate of wear and restrict access of air to the contact surfaces.

Contacts containing platinum metals (see PLATINUM GROUP METALS), such as palladium and rhodium, also may be subject to resistance problems from the formation of so-called frictional polymer (37) on the surfaces. This material is an insulating contaminant that originates in organic materials in the vicinity of the contact, such as organic vapors that adsorb on the surface, and subsequently polymerize to tough solids as a result of small mechanical displacements of the contacting surfaces. The thin coating of gold that usually is used on palladium and palladium alloy contact finishes significantly reduces any tendency to degrade by fretting.

Fiber Optic Connectors

Optical connectors are used to terminate and interconnect fiber optic cables (see FIBER OPTICS). Transmission of information by light through optical fibers made of glass or plastic is less expensive in many cases than transmission of electric signals through wire. An advantage of fiber optic technology is that it is

inherently immune from electromagnetic and radio frequency signal interference because no electrical signals are generated during lightwave data transmission. Fiber optics is popular in the telecommunication industry and there are expanding applications in the computer and aircraft and space industries.

Economic Aspects

Most electronic connectors are designed and made by one of the hundreds of companies devoted entirely to these products. The larger organizations both manufacture and sell internationally. In addition, this industry supports materials suppliers who provide piece parts such as metal stampings and molded connector bodies or offer a metal finishing service. The 10 largest connector companies have world sales ranging from about 300 to 2500 million U.S. dollars. These include AMP, MOLEX, AMPHENOL, 3M, ITT Cannon, Du Pont, Framatome, Hirose, Japan Aviation Electronics (JAE), and Japan Solderless Terminal (JST). Some large electrical connector users, such as those in the telecommunications, computer, and automotive fields, have captive connector manufacturing facilities and also offer products to the general trade.

The noncaptive worldwide market for electrical connectors, cable assemblies, and back panels was 19.1 billion dollars in 1992. Captive manufacture was about 31% in addition to this figure. The largest shipments were from North America, primarily the United States (38.6%); Japan (24.4%); Europe, mainly Germany, France, the UK, and Italy (25%); the Pacific Rim including Asia, mainly Taiwan, South Korea, Singapore, and Hong Kong (8.4%); and 3.6% for the rest of the world. The market is expected to increase slowly. Japan and Pacific Rim suppliers are growing in terms of their world share at the expense of North America and European suppliers (38).

BIBLIOGRAPHY

"Electrical Connectors" in *ECT* 3rd ed., Vol. 8, pp. 641–661, by M. Antler, Bell Telephone Laboratories.

1. M. Antler, W. G. Graddick, and H. G. Tompkins, *IEEE Trans. Parts, Hybrids, Packag.* **PHP-11**, 35 (1975).
2. R. S. Nelson, *Multewire Brush Contact Connector*, Report SAND 75-0245, Sandia Laboratories, Albuquerque, N.M.
3. P. Dobrogowski and J. Schroeder, III, *Proceedings of the Twelfth Annual Connector Symposium*, Electronic Connector Study Group, Cherry Hill, N.J., 1979, pp. 233–238.
4. R. F. Mallina, *Bell Syst. Tech. J.* **32**, 525 (1953).
5. J. O. Knudson, *Proceedings of the Sixth Annual Connector Symposium*, Electronic Connector Study Group, Cherry Hill, N.J., 1973, pp. 207–219.
6. H. B. Brown, *Proceedings of the Institute of Printed Circuits Conference*, Orlando, Fla. 1977.
7. H. J. Graff, J. M. Peacock, and J. J. Zalmans, *Bell Syst. Tech. J.* **40**, 131 (1963).
8. D. R. Frey and D. C. Borden, *Proceedings of the Second International Symposium on Subscriber Loops and Services*, London, May, 1976, pp. 68–72 (IEEE Conference Publication No. 37).
9. R. Holm, *Electric Contacts*, 4th ed., Springer-Verlag, New York, 1967, pp. 118–134.
10. Ref. 9, p. 8.
11. Ref. 9, pp. 9–26, 124, 125.
12. Ref. 9, pp. 87–92, 436–438.
13. Ref. 9, pp. 135–152.
14. M. Clarke, in R. Sard, H. Leidheiser, Jr., and F. Ogburn, eds., *Properties of Electrodeposits, Their Measurement and Significance*, The Electrochemical Society, Princeton, N.J., Chapt. 8, pp. 122–141.
15. M. Antler and J. J. Dunbar, *IEEE Trans. Components, Hybrids, Manufact. Technol.* **CHMT-1**(1), 17 (1978).
16. M. Antler, *IEEE Trans. Parts, Hybrids, Packag.* **PHP-9**, 4 (1973).
17. C. J. Thwaites, in F. H. Reid and W. Goldie, eds., *Gold Plating Technology*, Electrochemical Publications, Ltd., Ayr, Scotland, 1974, Chapt. 19, pp. 225–245.
18. M. Antler, in Ref. 17, Chapt. 36, pp. 478–494.
19. M. Antler and M. H. Drozdowicz, *Bell Syst. Tech. J.* **58**, 323 (1979).
20. M. Antler, M. H. Drozdowicz, and C. F. Hornig, *J. Electrochem. Soc.* **124**, 1069 (1977).
21. R. J. Russell, *Proceedings of the Seventh Annual Connector Symposium*, Electronic Connector Study Group, Inc., Cherry Hill, N.J., 1974, pp. 79–94.
22. F. E. Bader, *Proceedings of the 11th International Conference on Electric Contact Phenomena*, West Berlin, Germany, 1982, pp. 133–137.
23. A. Mendizza and P. C. Milner "Materials Technology," in *Physical Design of Electronic Systems*, Vol. II, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1970, Chapt. 5, p. 247.
24. T. P. Piccirillo and co-workers, *Proceedings of the Holm Seminar on Electrical Contacts*, Illinois Institute of Technology, Chicago, Ill., 1976, pp. 71–78.
25. C. Curry, *Proceedings of the Seventh Annual Connector Symposium*, Electronic Connector Study Group, Inc., Cherry Hill, N.J., 1973, pp. 211–216.
26. C. A. Haque, *Proceedings of the IEEE Holm Conference on Electrical Contacts*, Chicago, Ill., 1989, pp. 117–121.
27. M. Antler, *Proceedings of the IEEE Holm Conference on Electrical Contacts*, Boston, Mass., 1986, pp. 35–44.
28. J. R. Osias and J. H. Tripp, *Wear* **9**, 388 (1966).
29. I. F. Matthyse, *Basic Connection Principles*, 2nd ed., Burndy Corp., Norwalk, Conn., 1965.
30. R. W. Barnard and J. P. Pasternak, *Proceedings of the IEEE Holm Seminar on Electric Contact Phenomena*, Illinois Institute of Technology, Chicago, Ill., 1968, pp. 27–41.
31. J. E. Atkinson, *IEEE Trans. Reliab.*, 8 (June 1964).
32. *EEI-NEMA Standards for Connectors for Use Between Aluminum or Aluminum–Copper Overhead Conductors*, *EEI Pub. No. TDJ-162*, Edison Electric Institute, New York, Aug. 1973.
33. Ref. 23, p. 237.
34. Ref. 23, p. 249.
35. W. H. Abbott, *Proceedings of the IEEE Holm Conference on Electrical Contacts*, Chicago, Ill., 1987, pp. 62–78.
36. R. B. Waterhouse, *Fretting Corrosion*, Pergamon Press, New York, 1972, pp. 60–62.
37. M. Antler, *Wear* **106**(1–3), 5–33 (1986).
38. Technical data, *Fleck Report*, Fleck International, Huntington Beach, Calif., 1992.

General References

R. Holm, *Electric Contacts Handbook*, 4th ed., Springer-Verlag, New York, 1967; comprehensive treatment of electric contact theory.
Connectors and Interconnection Handbook, 5 vols., International Institute of Connector and Interconnection Technology, Inc., Anaheim, Calif.; emphasis on electronic connector design.

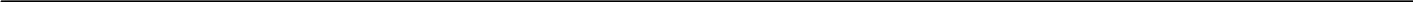
Comprehensive Bibliographies

Cumulative Index of the Proceedings of the IEEE Holm Conferences on Electrical Contacts and the International Conferences on Electrical Contacts, IEEE Service Center, Piscataway, N.J., IEEE Catalog Number JH 9412-8. 54 Conferences from 1953 through 1992.

Trade Publication

interConnection Technology, IHS Publishing Group, Libertyville, Ill.
Contributions to this paper were made by K. Fleck, M. Peel, and D. J. Williams.

Morton ANTLER
Contact Consultants, Inc.



ELECTRICAL INSULATORS.

See INSULATION, ELECTRIC.

ELECTRICALLY CONDUCTIVE POLYMERS

The realization of high electronic conductivities in organic polymers has developed to the state that commercial materials are now available. A significant research effort in chemistry, physics, and materials science has been devoted to developing this class of materials. The effect of chemical structure on electrical properties has been investigated in detail, and a general set of guidelines for the synthesis of an electrically conducting polymer has been established.

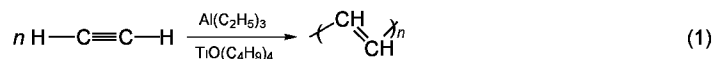
This article addresses the synthesis, properties, and applications of redox dopable electronically conducting polymers and presents an overview of the field, drawing on specific examples to illustrate general concepts. There have been a number of excellent review articles (1–13). Metal particle-filled polymers, where electrical conductivity is the result of percolation of conducting filler particles in an insulating matrix (14) and ionically conducting polymers, where charge-transport is the result of the motion of ions and is thus a problem of mass transport (15), are not discussed.

Generally speaking, electrically conductive polymers are composed of conjugated polymer chains with π -electrons delocalized along the backbone. In the neutral, or undoped, form the polymers are either insulating or semiconducting. The polymers are converted to the electrically conductive, or doped, forms via oxidation or reduction reactions which form delocalized charge carriers. Charge balance is accomplished by the incorporation of an oppositely charged counterion into the polymer matrix. The conductivity is electronic in nature and no concurrent ion motion occurs in the solid state. The redox doping processes are reversible and can be accomplished electrochemically. During electrochemical switching, ions do move into and out of the polymers as charge-balancing species for the charge carriers on the polymer backbone. Present and potential applications of conducting polymers utilize both their electronic (electrically conducting and optical) properties and their ionic properties, eg, as battery or sensor electrodes.

Synthesis of Electrically Conductive Polymers

A number of synthetic routes have been developed for the preparation of conjugated polymers. The diversity has been driven by the desire to examine many different types of conjugated polymers and attempts to improve material properties. Material property enhancement has centered on the synthesis of polymers that are processible in various forms. In some cases high molecular weight precursor polymers are prepared that can be converted into the conjugated form after processing. The conjugated polymer itself can also be soluble or fusible and is processed prior to doping. Five primary classes of conjugated polymers have been shown to exhibit high levels of electrical conductivity in the doped state. These include polyacetylenes, polyarylenes and polyheterocycles, poly(arylene vinylenes), and polyanilines. In addition, a number of multicomponent materials, usually polymer blends and composites, have been prepared in which at least one of the components is a conducting polymer.

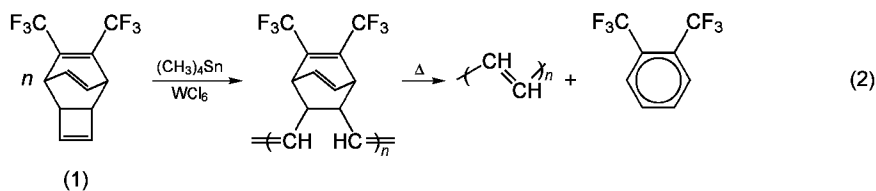
Polyacetylenes. The first report of the synthesis of a strong, flexible, free-standing film of the simplest conjugated polymer, polyacetylene [26571-64-2], $(\text{CH})_n$, was made in 1974 (16). The process, known as the Shirakawa technique, involves polymerization of acetylene on a thin-film coating of a heterogeneous Ziegler-Natta initiator system in a glass reactor, as shown in equation 1.



The resulting porous, fibrillar polyacetylene film is highly crystalline, so is therefore insoluble, infusible, and otherwise nonprocessable. It is also unstable in air in both the conducting and insulating form.

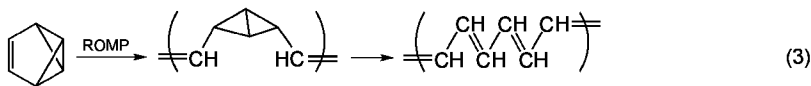
Much effort has been expended toward the improvement of the properties of polyacetylenes made by the direct polymerization of acetylene. Variation of the type of initiator systems (17-19), annealing or aging of the catalyst (20,21), and stretch orientation of the films (22,23) has resulted in increases in conductivity and improvement in the oxidative stability of the material. The improvement in properties is likely the result of a polymer with fewer defects.

Even with improvement in properties of polyacetylenes prepared from acetylene, the materials remained intractable. To avoid this problem, soluble precursor polymer methods for the production of polyacetylene have been developed. The most highly studied system utilizing this method, the Durham technique, is shown in equation 2.

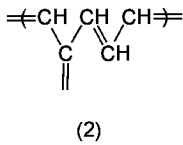


The method involves the metathesis polymerization of 7,8-bis(trifluoromethyl)tricyclo[4.2.2.0^{2,5}]deca-3,7,9-triene (1) (24,25), to form an acetone-soluble precursor polymer (26,27). After purification, thin films are cast from solution and the polymer is converted to polyacetylene by thermal treatment. This technique is useful only for the production of thin films because of the high exothermicity of the thermal conversion, which is a potential explosion hazard. However, the films produced in this manner are much less fibrillar and of higher density than the films made by the Shirakawa technique. Variable morphologies are also available using this method; cast films have a much higher amorphous content than Shirakawa films, but stretching of the precursor polymer before or during thermal treatment yields highly oriented and crystalline materials (28,29). Many systems, similar to the one shown in equation 2, have been studied. In addition to the readily thermally eliminated bis(trifluoromethyl) substituted benzene precursor polymers, other precursors with varying stabilities have been prepared (30).

A drawback to the Durham method for the synthesis of polyacetylene is the necessity of elimination of a relatively large molecule during conversion. This can be overcome by the inclusion of strained rings into the precursor polymer structure. This technique was developed in the investigation of the ring-opening metathesis polymerization (ROMP) of benzvalene as shown in equation 3 (31).

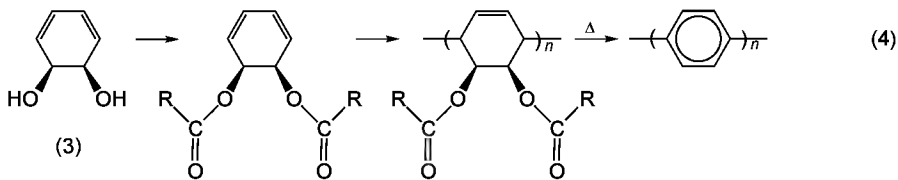


After the initial polymerization, the bicyclobutane rings in the polybenzvalene backbone can be isomerized to the conjugated system using transition-metal solutions, ie, HgCl₂, HgBr₂, and Ag⁺ salts. After conversion to (CH)_x, spectroscopic analysis indicates that the material has a high density of saturated defects, and low crystallinity. As with Durham (CH)_x, stretch orientation increases the extent of crystallinity of the polymer, but results in a more fibrillar morphology. Double bonds off the main chain, as shown in (2), and the presence of residual saturation, lead to a more highly disordered structure in polyacetylene synthesized in this manner when compared to Shirakawa (CH)_x.

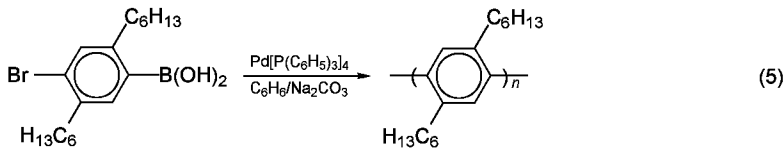


Copolymerizations of benzvalene with norbornene have been used to prepare block copolymers that are more stable and more soluble than the polybenzvalene (32). Upon conversion to (CH)_x, some phase separation of nonconverted polynorbornene occurs. Other copolymerizations of acetylene with a variety of monomers and carrier polymers have been employed in the preparation of soluble polyacetylenes. Direct copolymerization of acetylene with other monomers (33–39), and various techniques for grafting polyacetylene side chains onto solubilized carrier polymers (40–43), have been studied. In most cases, the resulting copolymers exhibit poorer electrical properties as solubility increases.

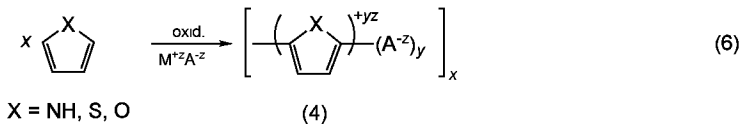
Polyphenylenes. Poly(*p*-phenylene) [25790-62-9] (PPP), synthesized using direct polymerization methods, yields oligomers of a completely intractable material (44–47). Although it is generally difficult to use intractable polymers in practical applications, sintering techniques are available which may make this polymer technologically useful. Electrochemical polymerization affords one route to produce intractable conducting polymers as useful films. Soluble precursor methods have been applied to PPP syntheses to circumvent this problem. Bacterial oxidation of benzene using the microorganism *Pseudomonas putida* to prepare 5,6-*cis*-dihydroxycyclohexa-1,3-diene (3), followed by esterification and free-radical polymerization, has made it possible to form a soluble PPP precursor as shown in equation 4 (48,49). A variety of esters (diacetate, dibenzoate, etc) have been investigated to yield high molecular weight, processible precursor polymers. Thermal conversion of thin films and fibers produce fully aromatic PPP. The *cis* diesters were also prepared using conventional chemical routes (50). The polymers were found to contain 10 to 15% 1,2 linkages rather than the desired all 1,4-linked structure.



In both studies described, the resulting PPP was completely intractable. Preparation of a soluble and processible form of PPP has been accomplished using dialkyl-substituted benzenes in the polymerization (51,52). Originally, Grignard coupling of alkyl-substituted 1,4-dibromobenzenes was employed, but low molecular weights were obtained because of loss of chain end functionality. Subsequently, an A–B step-growth polymerization technique was developed, as shown in equation 5. The resulting PPP had a significantly higher degree of polymerization.



Polyheterocycles. Heterocyclic monomers such as pyrrole and thiophene form fully conjugated polymers (4) with the potential for doped conductivity when polymerization occurs in the 2, 5 positions as shown in equation 6. The heterocycle monomers can be polymerized by an oxidative coupling mechanism, which can be initiated by either chemical or electrochemical means. Similar methods have been used to synthesize poly(*p*-phenylenes).



The electrochemical polymerization of pyrrole is generally believed to follow a radical step-growth mechanism. The process is illustrated in Figure 1. The monomer is oxidized at the anode to form radical cations, which quickly couple and eliminate two protons to rearomatize. The pyrrole dimer thus formed is more easily oxidized than pyrrole monomer, and is reoxidized to allow further coupling reactions to proceed. As the chain length of the growing oligomer increases, it becomes insoluble and deposits on the surface of the cell anode as a black film, where solid-state polymerization continues to occur. The polymer is further oxidized to a positively charged state (redox doped) and, to compensate for this charge, negatively charged ions from the electrolyte are incorporated into the film (4). Studies have indicated that coupling is primarily at the 2 and 5 positions, with small amounts at the 3 position which leads to structural disorder in the final polymer (54–60). Pyrrole can also be polymerized using chemical oxidants such as FeCl₃ in the presence of an electrolyte to form the conducting polymer in the form of a powder or coating.

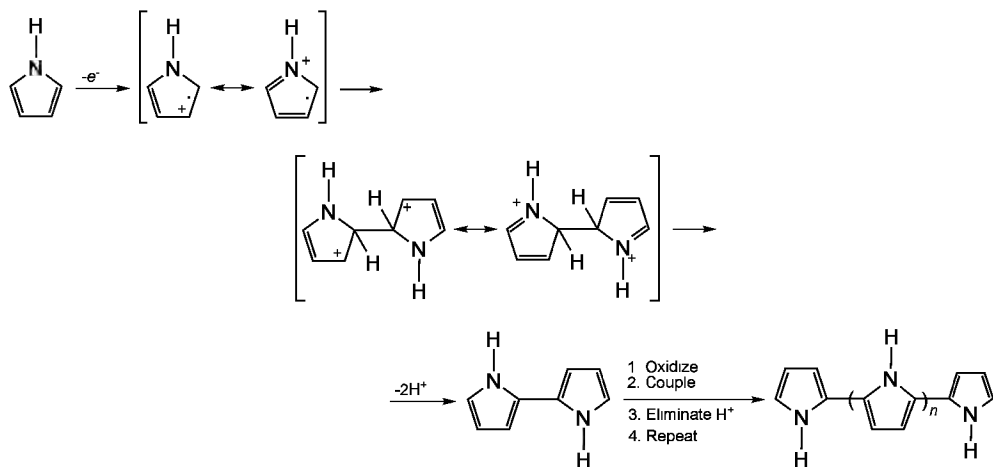
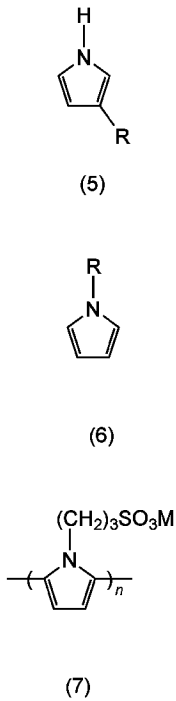
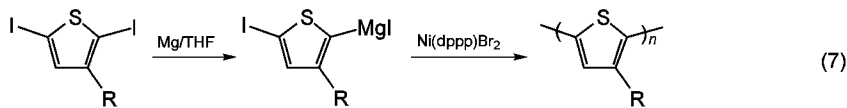


Fig. 1. Mechanism of electrochemical polymerization of pyrrole.

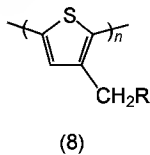
Significant variations in the properties of polypyrrole [30604-81-0] are controlled by the electrolyte used in the polymerization. Monoanionic, multianionic, and polyelectrolyte dopants have been studied extensively (61–67). Properties can also be controlled by polymerization of substituted pyrrole monomers, with substitution being at either the 3 position (5) (68–71) or on the nitrogen (6) (72–75). An interesting approach has been to substitute the monomer with a group terminated by an ion, which can then act as the dopant in the oxidized form of the polymer forming a so-called self-doped system such as the one shown in (7) (76–80).

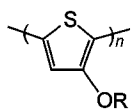


In all cases of electrochemically or chemically polymerized unsubstituted polypyrrole, the final polymer is intractable in both the conducting and insulating forms. In contrast, a broad number of substituted polythiophenes have been found to be processible both from solution and in the melt. The most studied of these systems are the poly(3-alkylthiophenes) (P3AT). Electrochemical synthesis of P3ATs was accomplished in 1986 (81). In the same year, a technique was reported for the chemical synthesis of P3ATs as shown in equation 7.

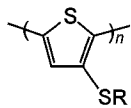


The synthesis involves the nickel-catalyzed coupling of the mono-Grignard reagent derived from 3-alkyl-2,5-diiodothiophene (82,83). Also in that year, transition-metal halides, ie, FeCl_3 , MoCl_5 , and RuCl_3 , were used for the chemical oxidative polymerization of 3-substituted thiophenes (84). Substantial decreases in conductivity were noted when branched side chains were present in the polymer structure (85). Polythiophenes with substituents other than alkyl groups at the 3 position have been prepared by the polymerization of substituted monomers. Many of these polymers have been substituted alkylthiophenes (8) where example side chains are $(\text{R})-\text{C}_6\text{H}_5$ (86–89), $-\text{OCH}_3$ (68), $-\text{NHC(O)}$ $(\text{CH}_2)_{10}\text{CH}_3$ (68), and $-\text{OSO}_2(\text{CH}_2)_3\text{CH}_3$ (90). Chiral side chains have also been employed (91,92). Poly(3-alkoxythiophenes) (9) (93–95) and poly(3-alkylthiothiophenes) (10) (95–98) have been prepared by both chemical and electrochemical methods.





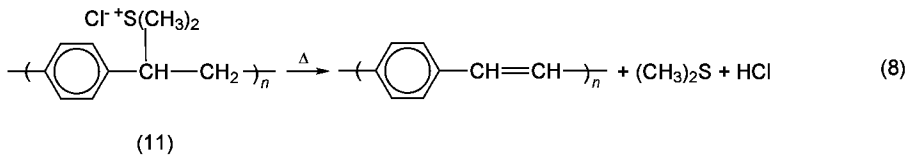
(9)



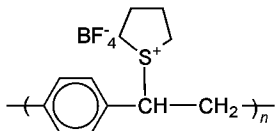
(10)

The properties of the final polymer (melting point, solubility, optical properties, etc) depend on the nature of the side chain incorporated into the structure, so that judicious choice of substituent can lead to material tailorability.

Poly(arylene vinylenes). The use of the soluble precursor route has been successful in the case of poly(arylene vinylenes), both those containing benzenoid and heteroaromatic species as the aryl groups. The simplest member of this family is poly(*p*-phenylene vinylene) [26009-24-5] (PPV). High molecular weight PPV is prepared via a soluble precursor route (99–105). The method involves the synthesis of the bis-sulfonium salt from *p*-dichloromethylbenzene, followed by a sodium hydroxide elimination polymerization reaction at 0°C to produce an aqueous solution of a polyelectrolyte precursor polymer (11). This polyelectrolyte is then processed into films, foams, and fibers, and converted to PPV thermally (eq. 8).



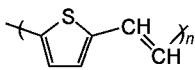
The nature of the sulfonium structure affects the yield and quality of the resulting PPV, and it has been found that use of cyclic sulfonium structures (12) is preferable (106). With cyclic sulfonium polyelectrolytes, more efficient elimination of sulfur and the counterion occurs during thermal conversion, so fewer *sp*³ defects are present in the final PPV.



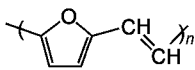
(12)

Substituted PPVs have been prepared using similar techniques. The earliest reports described methyl substituents (104,105), and more recently alkoxy substituents on the aromatic ring have been incorporated into the polymer structures (107–109). The advantage of long-chain alkoxy (butoxy or hexyloxy) substituents is that not only is the precursor polyelectrolyte soluble, but after conversion the substituted PPV is also soluble (110–112).

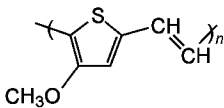
Heteroaromatic ring structures can also be incorporated into poly(arylene vinylene) structures using the same precursor polymer method shown for PPV. Poly(thienylene vinylene) (13) (113–118) and poly(furylene vinylene) (14) (119,120) have been prepared in this manner. In addition, alkoxy-substituted poly(thienylene vinylenes) (15) (119,121) have been synthesized. Various copolymers containing phenylene, thienylene, and furylene moieties have also been studied (120,122,123).



(13)

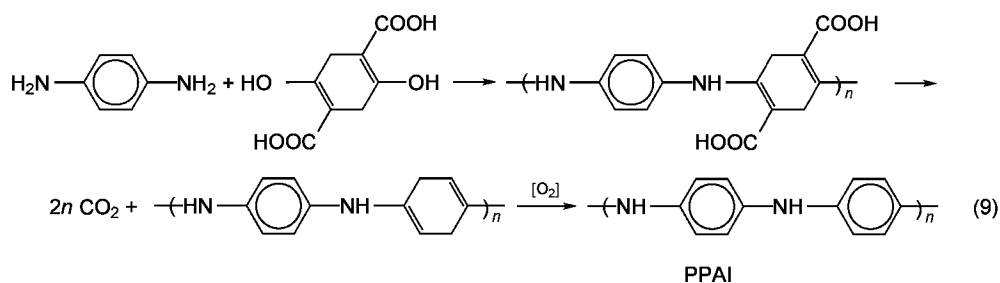


(14)

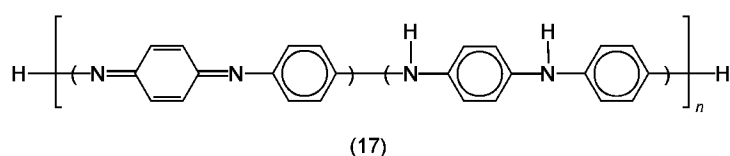
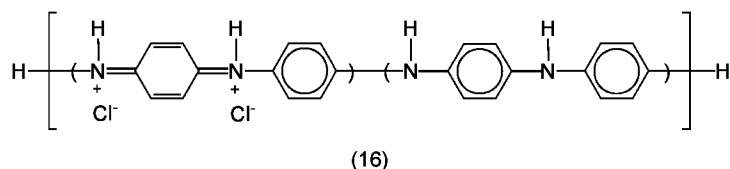


(15)

Polyanilines. Initial preparations of polyaniline (PANI) led to insoluble materials that were difficult to characterize. Use of model compounds and polymers (124,125) allowed for definitive structural analysis. Poly(*p*-phenylene amineimine) (PPAI) was synthesized directly to demonstrate that PANI is purely para-linked (126). The synthesis was designed so as to allow linkage through the nitrogen atoms only (eq. 9). Comparison of the properties of PPAI and PANI showed PPAI to be an excellent model both structurally and electronically.



PANI is more commonly prepared by polymerization of aniline using $(\text{NH}_4)_2\text{S}_2\text{O}_8$ in HCl (112,127). As prepared, it has structure (16) known as emeraldine hydrochloride. In this form, PANI is highly conductive but completely insoluble. When emeraldine hydrochloride is deprotonated with NH_4OH , the highly soluble emeraldine base (17) is produced. It is processible from organic solvents such as aqueous acetic acid or DMSO. It must then be treated with HCl to regenerate the again insoluble conducting form of the polymer.



As was the case for polythiophenes, substitution along the polyaniline [25233-30-1] backbone has been utilized as a means of improving processibility. Many derivatives are possible for PANI because of the possibility of substitution of the monomers at either the main-chain nitrogen atoms, or on the aromatic ring (128–135). Substituents studied have included alkyl, aryl, sulfonyl, and amino groups. The presence of a substituent dramatically decreases the yield of the polymerization and polymer molecular weight, the extent of the effect being proportional to substituent size. These substituted PANIs are insoluble as prepared, but treatment with base yields polymers that are highly soluble in common organic solvents.

Conducting Polymer Blends, Composites, and Colloids. Incorporation of conducting polymers into multicomponent systems allows the preparation of materials that are electroactive and also possess specific properties contributed by the other components. Dispersion of a conducting polymer into an insulating matrix can be accomplished as either a miscible or phase-separated blend, a heterogeneous composite, or a colloiddally dispersed latex. When the conductor is present in sufficiently high composition, electron transport is possible.

There are several approaches to the preparation of multicomponent materials, and the method utilized depends largely on the nature of the conductor used. In the case of polyacetylene blends, *in situ* polymerization of acetylene into a polymeric matrix has been a successful technique. A film of the matrix polymer is initially swelled in a solution of a typical Ziegler-Natta type initiator and, after washing, the impregnated swollen matrix is exposed to acetylene gas. Polymerization occurs as acetylene diffuses into the membrane. The composite material is then oxidatively doped to form a conductor. Low density polyethylene (136,137) and polybutadiene (138) have both been used in this manner.

Because of the aqueous solubility of polyelectrolyte precursor polymers, another method of polymer blend formation is possible. The precursor polymer is co-dissolved with a water-soluble matrix polymer, and films of the blend are cast. With heating, the fully conjugated conducting polymer is generated to form the composite film. This technique has been used for poly(arylene vinylenes) with a variety of water-soluble matrix polymers, including polyacrylamide, poly(ethylene oxide), polyvinylpyrrolidinone, methylcellulose, and hydroxypropylcellulose (139–141). These blends generally exhibit phase-separated morphologies.

The true thermoplastic nature of poly(3-alkylthiophenes), ie, solubility and fusibility, allows the use of compounding methods commonly used in the plastics industry for the preparation of composites of these polymers. The polymers can be co-dissolved with a matrix polymer, then processed from organic solution. Again the resulting blends are phase separated (142,143), but if the composition of conducting polymer is high enough, the conducting component forms the matrix with the insulating polymer dispersed within it, and high conductivity is possible. Melt processing has also been applied to poly(3-alkylthiophenes) (144–147). Films and sheets of blends of poly(3-octylthiophene) with polystyrene, polyethylene, poly(ethylene-*co*-butyl acrylate) and poly(ethylene-*co*-vinyl acetate) have been prepared using compression molding and film-blowing extrusion techniques.

Electrochemical polymerization of heterocycles is useful in the preparation of conducting composite materials. One technique employed involves the electro-polymerization of pyrrole into a swollen polymer previously deposited on the electrode surface (148–153). This method allows variation of the physical properties of the material by control of the amount of conducting polymer incorporated into the matrix film. If the matrix polymer is an ionomer such as Nafion (154–158) it contributes the dopant ion for the oxidized conducting polymer and acts as an effective medium for ion transport during electrochemical switching of the material.

Conducting polymer composites have also been formed by co-electrodeposition of matrix polymer during electrochemical polymerization. Because both components of the composite are deposited simultaneously, a homogenous film is obtained. This technique has been utilized for both neutral thermoplastics such as poly(vinyl chloride) (159), as well as for a large variety of polyelectrolytes (64–68, 159–165). When the matrix polymer is a polyelectrolyte, it serves as the dopant species for the conducting polymer, so there is an intimate mixing of the polymer chains and the system can be appropriately termed a molecular composite.

The preparation of molecular composites by electropolymerization of heterocycles in solution with polyelectrolytes is an extremely versatile technique, and many polyelectrolyte systems have been studied. The advantages of this method include the use of aqueous systems for the polymerization. Also, the physical and mechanical properties of the overall composite depend on the properties of the polyelectrolyte, so material tailorability is feasible by selection of a polyelectrolyte with desirable properties.

Polypyrrole has been used in the preparation of colloiddally dispersed latexes for processing into conducting polymer blends. Electrochemical polymerization of pyrrole in the presence of latex particles having a high concentration of bound sulfonate or sulfate groups on their surface results in the formation of thick heterogeneous films. These films are then ground, solvent swollen, and sheared to yield dispersions from which conducting films can be cast or spin coated (166). Chemical polymerization in aqueous solutions of methylcellulose also results in colloidal systems suitable for the casting of films of conducting polymer blends (167–169).

Redox Doping

In order to induce high electrical conductivity in organic conjugated polymers, charge carriers must be introduced. These charge carriers are created by removing electrons from, or adding electrons to, the delocalized π -electron network of the polymer, creating a conducting unit that is now a polymeric ion

rather than a neutral species. The charges introduced are compensated by ions from the reaction medium. This process is called doping by analogy to the changes that occur in inorganic semiconductors upon addition of small quantities of electronic defects. However, it proceeds through a different mechanism and is more precisely termed a redox reaction (7,170). The doping level, or ratio of charge carriers per polymer repeat unit, is generally between 0.2 and 0.4 in most polyarylenes (171) and can be determined by measuring the content of charge-balancing counterions. The ability to control the electrical properties of conducting polymers over wide ranges, by adjusting the redox doping level, has created interest in these materials for a number of emerging applications (10).

Charge Carriers in Conducting Polymers. Metals have unpaired electrons and their highest occupied electronic levels are half-occupied. Electrical conductivity results from the fact that the electrons can move readily under an electric field since there is no forbidden gap between the highest occupied and lowest unoccupied electronic levels. Conductivity is limited, therefore, by defects in the lattice and vibrational distortions (also called phonons). Since this phonon activity increases with elevated temperatures, the conductivity of metals increases as the temperature is decreased. However, the electrons in conjugated organic polymers, as in inorganic semiconductors, are paired, creating a gap between the highest occupied levels (the valence band) and the lowest unoccupied levels (the conduction band). The energy difference between these bands gives rise to the intrinsic insulating or semiconducting properties of conjugated organic polymers. The moderate conductivity of these materials is a result of thermal excitation of valence electrons into the conduction band. Therefore, the conductivity of semiconductors and conjugated organic polymers increases with increasing temperature in the neutral or undoped state.

The mechanism for the conductivity increase resulting from doping in inorganic semiconductors involves the formation of unfilled electronic bands. Electrons are removed from the top of the valence band during oxidation, called *p*-type doping, or added to the bottom of the conduction band during reduction, termed *n*-type doping. Extension of this argument to the case of conjugated organic polymers was found to be inaccurate as the conductivity in many conducting polymers was found to be associated with spinless charge carriers. *In situ* epr electrochemistry techniques have shown that the conducting entity in polyacetylene (172), polypyrrole (173), polythiophene (174,175), and poly(*p*-phenylene) (176) can be spinless, although evidence exists for mixed valence charge carriers as well (177).

The conductivity increase following doping in conjugated polymers is explained in terms of local lattice distortions and localized electronic states. In this case the valence band remains full and the conduction band remains empty so that there is no appearance of metallic character. When the polymer chain is redox doped, a lattice distortion results and the equilibrium geometry for the doped state is different than the ground state geometry. This is evident in many small organic molecules, for example the ground state geometry of biphenyl is benzenoid but the geometry of its radical cation is quinoid (178).

Many conjugated polymers have nondegenerate ground states which behave similarly upon redox doping. When one electron is removed (or added) to the polymer chain a radical cation (or anion) is formed. This results in a lattice distortion which leads to an upward shift of the highest occupied molecular orbital (HOMO) and a downward shift in the lowest unoccupied molecular orbital (LUMO). Although the radical ion is expected to be delocalized over the entire polymer chain, the species is localized, with a localized lattice distortion creating a localized electronic state. Since the geometry of the chain between the radical and the ion must be distorted, and the energy of the distorted geometry is normally higher than that of the ground state, separation and delocalization of the radical ion results in energetically unfavorable further lattice distortions. This radical ion associated with a lattice distortion is called a polaron (179) as shown in Figure 2 for PPP. A similar argument can be used when a second electron is removed from (or added to) this site. The resulting species is a dication (or dianion), with a lattice distortion separating the two charges. This species is termed a bipolaron and is also illustrated in Figure 2.

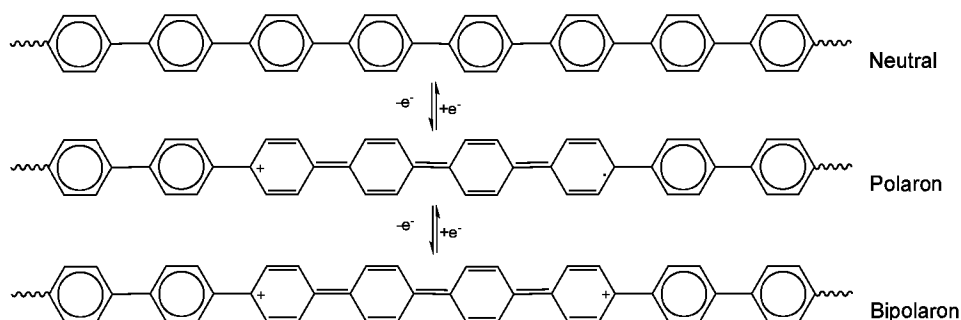


Fig. 2. Lattice distortions associated with the neutral, polaron, and bipolaron states in poly(*p*-phenylene).

Although this model can be applied to most conducting polymers, polyacetylene is a special case because of the degeneracy of its ground state. The energy of the distorted geometry, where the single and double bonds are simply reversed, is equivalent to the energy of the initial structure (Fig. 3). This fact allows the charges formed during the doping of polyacetylene to readily separate because there is no increase in distortion energy. This type of charge carrier is called a soliton and is unique to conjugated polymers, such as polyacetylene, which have degenerate ground states (11,180).

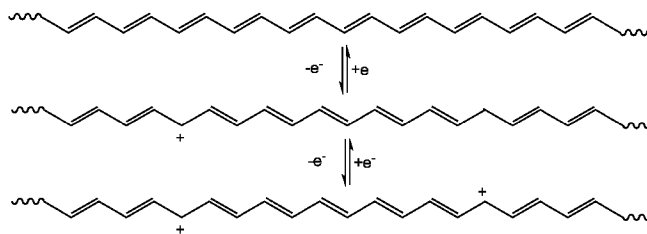


Fig. 3. Lattice distortions associated with the neutral and soliton states in polyacetylene.

Doping Processes. Redox doping can be accomplished through both chemical and electrochemical means. Vapor-phase doping of neutral polyacetylene has been carried out using AsF_5 and I_2 as oxidants. During this process the dopant is reduced to form a charge-balancing ion such as AsF_6^- and I_3^- . Polypyrrole can be doped with oxygen because of the low oxidation potential of the neutral polymer (11). Solution doping of many conjugated polymers in suspension or, in the case of processible derivatives, in solution, has been accomplished using a variety of oxidizing and reducing agents. Typical chemical dopants include FeCl_3 , NOPF (181), and sodium naphthalide. Conducting polymers that are synthesized using oxidative coupling techniques are obtained in the oxidized form.

These conjugated polymers can be chemically and electrochemically reduced and reoxidized in a reversible manner. In all cases the charges on the polymer backbone must be compensated by ions from the reaction medium which are then incorporated into the polymer lattice. The rate of the doping process is dependent on the mobility of these charge compensating ions into and out of the polymer matrix.

Electrogenerated conducting polymer films incorporate ions from the electrolyte medium for charge compensation (182). Electrochemical cycling in an electrolyte solution results in sequential doping and undoping of the polymer film. In the case of a *p*-doped polymer, oxidation of the film results in the

formation of cations on the polymer backbone and the introduction of anions into the film. Since the ions that are transported into the polymer matrix generally make up 20–40% of the doped polymer, the choice of electrolyte has a marked influence on the mechanical and electrochemical properties of the film. Tetraalkylammonium salts are often utilized because of their high solubility in electrochemical solvents and their large electrochemical window. Their perchlorate, fluoroborate, and hexafluorophosphate salts are commonly used in the electrochemical doping of polythiophenes and polypyrroles. However, superior film-forming properties have been observed in the electrosynthesis and doping of polypyrrole using organic sulfonates, such as toluenesulfonate and other arylsulfonates (183).

In the special case when polyelectrolytes (184) or multivalent anions (185) are used as the electrolyte during electrochemical polymerization, the dopant anion is entrapped because of its large size and multiple charges. As such, it cannot be transported into and out of the films. Cations from the electrolyte medium move into the film during electrochemical reduction to compensate the negative charges on the polyelectrolyte when the electroactive polymer is in its neutral form. Conversely, when the electroactive polymer is oxidized, the cations are expelled into the electrolyte medium and the negative charges on the polyelectrolyte are free to compensate the newly formed positive charges on the conducting polymer backbone.

Optical Properties. The energy difference between the valence band and the conduction band in conjugated polymers is referred to as the band gap and can be determined using optical spectroscopy. In the neutral (or insulating) form, conducting polymers exhibit a single electronic absorption in either the visible or ultraviolet region attributed to the electronic transition between the HOMO and LUMO levels as shown in Figure 4. The band gap is generally determined from the onset of this transition. This figure shows a band diagram of a conjugated polymer with a nondegenerate ground state, and the possible transitions in the neutral, polaron, and bipolaron states. Table 1 displays optically determined band gaps for several conducting polymers.

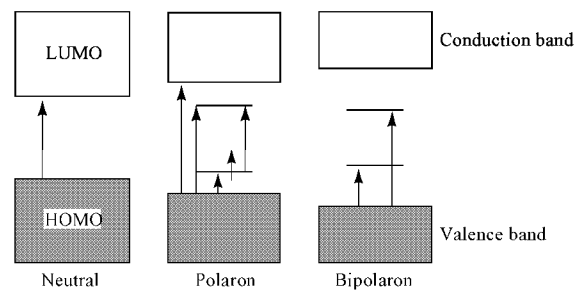


Fig. 4. Band diagram of a nondegenerate ground state conducting polymer.

Table 1. Optical Band Gaps For Conjugated Polymers

Polymer	Band gap, eV	Reference
polyacetylene	1.4	11
polypyrrole	2.5	8
polythiophene	2.0	186
poly(3-methylthiophene)	2.2	187
polyfuran	2.7	188
poly(3-hexylfuran)	2.5	188,189
poly(<i>p</i> -phenylene vinylene)	2.4	190
poly(<i>p</i> -phenylene)	3.0	191
poly(isothianaphthene)	1.0	192

In the doped form, transitions between the band edges and newly formed intragap electronic states are observed in the optical spectra of conducting polymers. When polarons are present as charge carriers, an additional transition is apparent which corresponds to the electronic transitions between the two gap states. Since the intragap electronic states are taken from the band edges, the band gap increases with increasing doping levels. Also, since a bipolaron creates a larger lattice distortion than a polaron, the gap states are further away from the band edges in the bipolaron model.

The changes in the optical absorption spectra of conducting polymers can be monitored using optoelectrochemical techniques. The optical spectrum of a thin polymer film, mounted on a transparent electrode, such as indium tin oxide (ITO) coated glass, is recorded. The cell is fitted with a counter and reference electrode so that the potential at the polymer-coated electrode can be controlled electrochemically. The absorption spectrum is recorded as a function of electrode potential, and the evolution of the polymer's band structure can be observed as it changes from insulating to conducting (11).

An optoelectrochemical spectrum of polypyrrole is shown in Figure 5 (193). The spectrum of the neutral polymer film contains a single high energy absorption which peaks at 3.2 eV and corresponds to the band-gap transition. At moderate doping levels, absorptions at 0.7 and 2.1 eV are also visible which correspond to the transitions between the intragap states and the band edges along with an absorption at 1.4 eV resulting from the transition between the two gap states. This indicates the presence of polaronic charge carriers. At higher doping levels, the absorption attributed to the intragap transition disappears and the band-gap absorption diminishes significantly and increases in energy as expected.

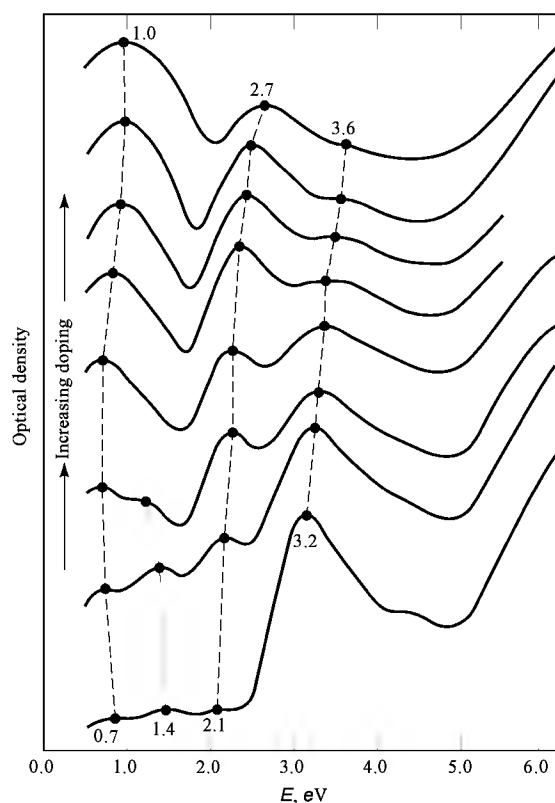


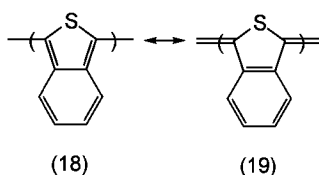
Fig. 5. Evolution of optical spectra of polypyrrole during electrochemical doping.

Structural Influence on Optical Properties. Since the formation of these types of charge carriers is essential for electrical conductivity in conjugated organic polymers, an important factor in the structural design of conducting polymers is the ease with which they can be oxidized or reduced. The ionization potential of a class of conducting polymers can be altered by modification of the chemical structure. Substitution of the polymer with electron-donating groups has been shown to lower the ionization potential for *p*-type doping. The increased electron density along the conjugated system allows for easier removal of electrons during oxidation.

Another important factor affecting the electronic properties is the steric barrier to planarity along the polymer chain. Since polyheterocycles and polyarylenes must adopt a planar geometry in the ionized state to form quinoid-like segments, steric factors that limit the ability of the polymer to adopt geometries which are planar with respect to adjacent rings have a detrimental effect on the electronic properties (181).

In the case of polythiophenes, for example, the band gaps of alkoxy-substituted derivatives, substituted in either the 3 position (194) or both the 3 and 4 positions (195), are lower than that of the parent polymer, however the band gaps of 3-alkyl derivatives are slightly higher than polythiophene (187). The electron-donating ability of the alkoxy groups are able to outweigh the steric interactions they create. Conversely, the steric factors involved with the alkyl groups at the 3 position of the thiophene ring create geometric strains that outweigh the contribution of these substituents to the electron density of the polymer chain.

The structural effects on the band gap of conjugated organic polymers have been used to create materials with unique optical properties. Design of low band-gap polymers has centered on decreasing the energy difference between the ground state and distorted geometries. One example of such a polymer is poly(isothianaphthene) (18) (192). The six-membered ring fused to the thiophene backbone becomes aromatic when the thiophene moiety assumes a quinoid-like geometry (19). This significantly reduces the energy difference between these two geometries and yields a polymer with a band gap of 1.0 eV.



When doped, low band-gap polymers have optical transitions in the infrared region of the spectrum, and therefore transmit more visible light in the conducting form than in the insulating form. This feature enables this class of conducting polymers to be investigated for a number of optical applications where both electrical conductivity and optical transparency are desired.

Electrical Properties

The electrical conductivity, σ , of a material is equal to the inverse of its specific resistivity, ρ , and is a measure of a material's ability to transport electrical charge. It is determined by measuring the resistance, R , to charge transport through a known volume as shown below where L is the length over which the resistance is being measured and A is the cross-sectional area through which the current is passing.

$$\sigma = 1 / \rho = L / RA$$

Thus electrical conductivity is commonly measured in units of $S \cdot cm$ ($\Omega^{-1} cm^{-1}$). Various experimental methods have been used to measure the electrical conductivity of conductive polymers, eg, 4-probe method, Van der Pauw method, etc, and are well documented in the literature.

In simplistic terms, the conductivity of a material is controlled by both the density, n , and mobility, μ , of the charge carriers, having a charge of e .

$$\sigma = ne\mu$$

One of the benefits of conductive polymers is that the number of charge carriers can be quite high because of the high density of redox sites along the conjugated backbones. The mobility of the charges are controlled by both intrachain and interchain interactions. Defect-free conjugated polymers that are highly oriented in the direction of electrical transport, can exhibit high mobilities along the chains. At the same time, strong π -overlap interchain interactions allow charge carriers to "hop" from chain to chain and ultimately high electrical conductivities are possible.

Table 2 shows the present state-of-the-art for the electrical conductivity of doped conjugated polymers. The magnitude of the electrical conductivity in polymers is a complex property determined by many structural aspects of the system. These include main-chain structure and π -overlap, molecular

weight and polydispersity, charge carrier density determined by redox doping, interchain interactions controlled by both main-chain and side-chain structures, and chain orientation which can be affected by sample processing conditions. In addition, sample morphology plays a significant role in limiting or assisting in charge-transport through a solid. Considerations here include crystalline amorphous material content and bulk dense morphologies, as compared with highly open fibrillar morphologies. Finally, stability of the doped polymer complexes must be accounted for. The conductivities listed are those that are reproducible in various laboratories and are not the maximum conductivities reported.

Table 2. Electrical Conductivity Ranges for Conductive Polymers

Polymer	S cm
polyacetylene	10 ³ –10 ⁵
poly(<i>p</i> -phenylene)	10 ³
polyazulene	10 ⁰
poly(<i>p</i> -phenylene sulfide)	10 ¹ –10 ²
poly(<i>p</i> -phenylene vinylene)	10 ³ –10 ⁴
poly(thienylene vinylene)	10 ³
polythiophene	10 ²
poly(3-alkyl thiophene)	10 ² –10 ³
polyisothianaphthene	10 ¹
polypyrrole	10 ² –10 ³
polyfuran	10 ²
polyaniline	10 ² –10 ³

With these many considerations, reported values of electrical conductivity vary widely as a function of the method of preparation and handling. As research in the field has progressed, better synthetic techniques and methods of processing samples have been developed, yielding polymers with higher electrical conductivities. Two examples illustrate this point well. Polyacetylene prepared using the Shirakawa method (16) in the late 1970s typically exhibited maximum conductivities of 500–1000 S cm. These Ziegler-Natta polymerization methods have been refined such that polyacetylenes having conductivities of 20,000–30,000 S cm are reproducibly prepared (196) and values higher than 100,000 S cm have been reported. These high conductivities are only slightly lower than those for many metals ($\sigma_{\text{copper}} = 600,000 \text{ S cm}$) and when density differences are accounted for ($d_{\text{copper}} = 8.9 \text{ g cm}^{-3}$; $d_{\text{polyacetylene}} = 1.1 \text{ g cm}^{-3}$), their conductivities are essentially equivalent on a weight basis.

Poly(*p*-phenylene vinylene) prepared by coupling polymerization has a low degree of polymerization and is obtained as an insoluble, relatively amorphous powder. When doped, pressed pellets of these polymers have conductivities of 1–10 S cm. When poly(phenylene vinylene) is prepared by a soluble precursor polymer, strong continuous films can be cast that can be stretch aligned to yield oriented films with a high degree of order. Doping these films using similar conditions to those for the powders yield materials with conductivities of 2500–3000 S cm.

Stability of Conducting Polymers

The rapid development of many processible conducting polymers that can now be obtained as films, fibers, and molded shapes, suggests that they will be quite useful in a number of applications. One of the concerns presently being addressed is the stability of the conductive polymers during these specific uses.

Although polyacetylene has served as an excellent prototype for understanding the chemistry and physics of electrical conductivity in organic polymers, its instability in both the neutral and doped forms precludes any useful application. In contrast to polyacetylene, both polyaniline and polypyrrole are significantly more stable as electrical conductors. When addressing polymer stability it is necessary to know the environmental conditions to which it will be exposed; these conditions can vary quite widely. For example, many of the electrode applications require long-term chemical and electrochemical stability at room temperature while the polymer is immersed in electrolyte. Aerospace applications, on the other hand, can have quite severe stability restrictions with testing carried out at elevated temperatures and humidities.

It is well known that the doped polypyrroles exhibit one of the best stabilities of electrical conductivity. Typically, poly(pyrrole tosylate) undergoes about a 10% conductivity decrease in a year at room temperature in air. The stability of the conductivity of polypyrrole and a series of poly(3-alkylthiophenes) has been compared in detail (197,198) under systematically controlled environmental and thermal conditions. When poly(pyrrole tosylate) is exposed to a dry N₂ atmosphere at 120°C there is no evident conductivity change for at least 24 hours. Poly(3-alkyl-thiophene tetrachloroferrates), on the other hand, show distinct conductivity decreases because of a thermal undoping process (199). The conductivity change is a function of the length of the pendent alkyl chain, with higher stability exhibited by polymers having the shortest alkyl chains. For example, poly(butylthiophene) exhibits a one-half order of magnitude drop in conductivity in the same amount of time that the conductivity of poly(3-octylthiophene) drops two orders of magnitude.

Poly(pyrrole tosylate) exhibits a slightly lower stability in air at 120°C with conductivity changes on the order of 10% in 24 hours, as a result of reactions with environmental O₂ and H₂O. To circumvent this, a study of the mechanical integrity and thermal stability of polypyrroles containing a broad range of dopant ions (200) showed that encapsulation of the conducting polymer in an epoxy matrix led to a great improvement in stability. This is illustrated in Figure 6 where an electrochemically prepared poly(pyrrole tosylate) film and a polypyrrole-coated textile were encapsulated and their electrical resistivities monitored during exposure to 80°C and 95% relative humidity (rh) for one month. After an initial conductivity change, the film maintains a high conductivity throughout the entire time. The polypyrrole-coated textile, although exhibiting no change in conductivity over the entire experiment, has a higher overall resistivity than the film.

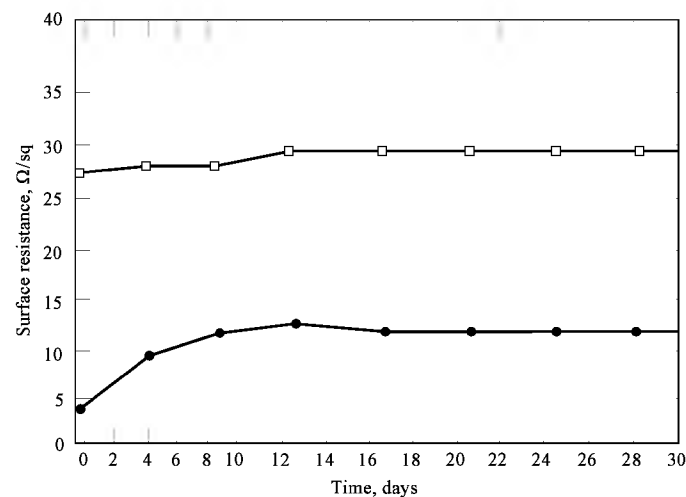


Fig. 6. Thermal stability of epoxy-encapsulated poly(pyrrole tosylate) film, ●, 0.5 Ω sq, and polypyrrole-coated textiles, □, 20 Ω sq, with exposure to 80°C and 95% rh.

One of the more stable of the electronically conducting polymers is polyaniline (PANI) which is easily prepared from the oxidative polymerization of aniline in solution. This process is now being used commercially by Allied-Signal, Inc. to prepare an intrinsically conductive polymer powder (Versicon). Electrical conductivities of 2–4 S cm are exhibited for materials having particle sizes of 3–100 μm. This powder can be either dispersed or blended with typical thermoplastic polymers in order to be processed. This form of PANI can withstand processing temperatures of up to 240°C and is specified for use for long periods of time (yr) at 100°C. In conjunction with Americhem Inc. and Zippering Kessler & Co., Allied-Signal is marketing blends of PANI with poly(vinyl chloride) for compression molding, injection molding, and extrusion. These blends can be processed at temperatures between 135 and 150°C and have potential applications in the electronic shielding and packaging industry.

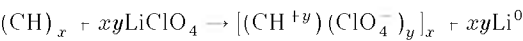
Applications of Conducting Polymers

The novel combination of electrical, electrochemical, and chemical properties of conductive polymers may lead them to be used in a number of applications. At this time (1993), the systems do not have appropriate electronic properties and stabilities to use conditions to be considered as replacement materials for metals or semiconductors. Of course, this does not preclude further advances from being made in the near future. Because of the present advanced commercial development of Versicon by Allied-Signal, Inc. relative to the other conducting polymers, it is being examined for many applications. Some of those suggested by the manufacturer include lightly colored coatings for antistatic films, and blends for electrostatic dissipation, and as corrosion-preventive paints. Conductive hosings, gaskets, cable shields, and radar-absorbing coatings are being tested for emi shielding applications. The conductive properties are being exploited in adhesives, inks, electrodes, and resistive heaters. In electronics applications PANI has been used as a discharge layer for electron-beam lithography (201), as a removable sem discharge layer (202), and as a conducting electrode for electrolytic metallization of copper on through-holes (203).

Many of the applications of conductive polymers utilize their unique properties and advantages over other material systems, for example low density and controllable electrical properties. The following examples demonstrate the versatility of conducting polymers in technology.

Charge-Storage Batteries. The reversible redox chemistry of conductive polymers has allowed the materials to be used as electrodes in rechargeable storage batteries. Initially it was hoped that the low density of polymer electrodes would yield lightweight high energy density batteries. This has not been realized because the polymers have low charge densities relative to metal electrodes. The charging and discharging reactions (redox doping and undoping) create charged carriers along the conjugated backbones with concurrent insertion or emission of charge-balancing counterions from the electrolyte. As such, there is no deposition or dissolution of the electrode material and it can be dimensionally stable. This is one limitation of metal electrodes in battery cells where repeated deposition and dissolution cause morphological features at the electrode surface which can separate and break from the electrode. In some instances, high surface area dendrites can form on metal surfaces, ultimately leading to an electrical contact between the two electrodes.

Early research, in which conductive polymers were considered as potential rechargeable battery electrode materials, focused on polyacetylene (204). Polyacetylene: lithium cells were assembled in organic electrolytes such as LiClO₄ in propylene carbonate. When the polyacetylene is in the fully doped state the battery is charged and exhibits an open-circuit potential (*V*_{oc}) of 3.3–3.7 volts. Spontaneous discharge occurs when these two electrodes are shorted together with undoping of the polyacetylene occurring as shown by the overall reaction in the following equation.



This reaction is fully reversible with application of a sufficiently positive potential to the polyacetylene which recharges the cell. Conductive polymer batteries were developed further by VARTA BASF who investigated a series of polypyrrole-based cells (205). Table 3 compares the results of conductive polymer-based cells with typical battery systems. It can be seen that the *V*_{oc}s and maximum energy densities are in useful ranges. Further work at Allied-Signal (206) has shown that a number of conducting polymers can be used in composite electrodes for the design of high energy density batteries. This work on AF- and AA-sized metal cans demonstrates the usefulness of conducting polymer battery technology with cells showing enhanced properties relative to commercially available Ni Cd cells.

Table 3. Battery Characteristics for Conductive Polymer Electrodes

Cell composition	<i>V</i> _{oc} ^a	Maximum energy density, Wh/kg
polypyrrole ClO ₄ Li ⁰	3.3	300
polyacetylene ClO ₄ Li ⁰	3.3	300
polyaniline BF ₄ Li ⁰	3.5	370
polyaniline BF ₄ Li ⁰	3.0	160
polyacetylene Li ⁺ Li ⁰	1.5	345
NiOOH Cd	1.3	209

^a Open-circuit potential.

The first true commercialization of a conductive polymer electrode in a battery cell has been carried out by Bridgestone Sienko who market a

polyaniline:lithium coin-type, or button, cell (207). These cells exhibit near 100% coulombic efficiencies up to cell voltages of 3.5 volts, with recommended use voltages of 3.0 volts. At this cell voltage the battery exhibits discharge characteristics of 3.0 mAh or 60 A·h/kg. These cells have cycle lives that are a function of discharge depth. At low discharge depths (<10%) the battery could be recharged up to 10,000 times. At 33% discharge the cell could be recharged 1000 times and at a full 100% discharge with each cycle, the cell could be recharged 200 times. These cells proved to be quite stable with storage. Storage for 120 days at 60°C, which can be correlated with five years storage under ambient conditions, led to only a 15% loss of initial discharge capacity. After storage at room temperature for three months the cells retained more than 80% of their charge capacity.

Conductive Textiles. The surface-confined chemical oxidative polymerization of pyrrole and aniline has been used by Milliken Corp. (Spartanburg, South Carolina) to produce textile materials with electrically conductive polymer coatings (208,209). Exposure of various fabrics, including nylons, polyester, and quartz, to an appropriate oxidant leads to a preferential surface adsorption. Subsequent addition of dopant anion and monomer leads to polymerization on the fiber surfaces and formation of coherent, well adhered, conductive polymer coatings about 1 μm in thickness. Interestingly, the nature of the fiber was found to have little effect on the process and thus it could be extended to many textile materials. Electrical properties were measured as surface resistivities which exhibited lower limits (highest conductivities) of 10 Ω/sq; polypyrrole coatings are about one order of magnitude more conductive than polyaniline coatings. Environmental aging of the conductive textiles with polypyrrole coatings, carried out at 71°C and 50% relative humidity, showed conductivity decreases of ca 50% in 20–30 days. Encapsulation of these textiles into epoxy laminates, as described earlier for polypyrrole films, improved stability considerably. No conductivity changes were evident after exposure to 71°C and 95% relative humidity for over one month.

Chemical and Biochemical Sensors. The sensitivity of the electrical properties of conductive polymers to chemical stimuli suggests they may prove useful in a number of sensing applications.

One early program carried out at Allied-Signal, Inc. proposed the use of conductive polymers in remotely readable indicators (210). Conductivity changes induced in the conductive polymer could be read externally and the history of the sample known. Systems designed to detect time–temperature, temperature limit, radiation dosage, mechanical abuse, and chemical exposure were developed.

Conductive polymers can change their electrical conductivities by many orders of magnitude when exposed to specific gaseous reagents (211). The reactivity of polypyrrole to NH₃, NO₂, and H₂S was tested to examine this phenomena. These sensors proved to be quite sensitive with large resistance changes occurring with exposure to an atmosphere containing 0.01–0.1% of the reactive gas. In all three cases the reactions were shown to be reversible as removal of the gas from the atmosphere led to recovery of the original conductivity.

Entrapment of biochemically reactive molecules into conductive polymer substrates is being used to develop electrochemical biosensors (212). This has proven especially useful for the incorporation of enzymes that retain their specific chemical reactivity. Electropolymerization of pyrrole in an aqueous solution containing glucose oxidase (GO) leads to a polypyrrole in which the GO enzyme is co-deposited with the polymer. These polymer-entrapped GO electrodes have been used as glucose sensors. A direct relationship is seen between the electrode response and the glucose concentration in the solution which was analyzed with a typical measurement taking between 20 to 40 s.

Electromagnetic Shielding and Antistatic Coatings. One of the most likely applications for which the properties of electrically conductive polymers will be exploited is in the general area of electromagnetic shielding (213) and antistatic coatings. The processibility of the polyanilines and the poly(3-alkylthiophenes), and the conductive polymer-coated textiles described above, has allowed them to be prepared in forms applicable to typical industrial processes for these applications. In the case of the polyanilines, they have been blended with PVC to form the Incoblend family of products. A number of conducting polymer blends have been shown to be efficient in emi-shielding applications using both far-field and near-field shielding measurements (213). Commercial specifications can be obtained on specific materials from Allied-Signal (Morristown, New Jersey); Americhem, Inc. (Cuyahoga Falls, Ohio), polyanilines; Uniax, Inc. (Santa Barbara, California), poly(3-alkylthiophenes); and Milliken Corp. (Spartanburg, South Carolina), textiles.

Gas Separation Membranes. Studies suggest that, because of the fixed ionic sites along the polymer backbones, conducting polymers may prove useful as gas separation membranes (214,215). The ability to switch the polymers between charge states (doped and undoped) suggests that external control of gas permeability and selectivity may be possible. Studies of PANI (215) showed it served as an excellent membrane material for the separation of the hydrogen–nitrogen, oxygen–nitrogen, and carbon dioxide–nitrogen gas pairs with extremely high selectivities. By utilizing dopants of varied sizes the separation properties of the membranes to different gases can be controlled.

Electrooptics and Electrochromism. During electrochemical switching, distinct changes occur in the optical properties (color and extinction coefficient) of conducting polymer films. This has led many investigators to examine the feasibility for use in electrically switchable, electrochromic or smart windows. Possibly the most intriguing of the polymers to be employed here are those with small electronic band gaps. As exemplified by polyisothianaphthene (192), thin polymer films are relatively transparent when held in the conducting state and opaque in the insulating state. Using the concepts originally put forth for polyisothianaphthene, a number of low band-gap polymers have been synthesized with similar properties.

The application of a high electric field across a thin conjugated polymer film has shown the materials to be electroluminescent (216–218). Until recently the development of electroluminescent displays has been confined to the use of inorganic semiconductors and a limited number of small molecule dyes as the emitter materials. Expansion to the broad array of conjugated polymers available gives advantages in control of emission frequency (color) and facility in device fabrication as a result of the ease of processibility of soluble polymers (see CHROMOGENIC MATERIALS, ELECTROCHROMIC).

BIBLIOGRAPHY

"Polymers, Conductive" in *ECT* 3rd ed., Vol. 18, pp. 755–793, by C. B. Duke and H. W. Gibson, Xerox Corp.

1. A. G. MacDiarmid and A. J. Heeger, *Synth. Met.* **1**, 1013 (1978).
2. G. B. Street and T. C. Clarke, *IBM J. Res. Dev.* **25**, 51 (1981).
3. G. Wegner, *Angew. Chem. Int. Ed. Engl.* **20**, 361 (1981).
4. K. J. Wynne and G. B. Street, *IEEC Prod. Res. Dev.* **21**, 23 (1982).
5. R. H. Baughman, *Contemp. Topics Polym. Sci.* **5**, 321 (1984).
6. R. L. Greene and G. B. Street, *Science* **226**, 651 (1984).
7. J. L. Bredas and G. B. Street, *Acc. Chem. Res.* **18**, 309 (1985).
8. J. R. Reynolds, *J. Molec. Elec.* **2**, 1 (1986).
9. A. J. Epstein, in T. A. Skotheim, ed., *Handbook of Conducting Polymers*, Vol. 2, Marcel Dekker, New York, 1986, p. 1041.
10. R. S. Potember and co-workers, *Polymer* **28**, 574 (1987).
11. A. O. Patil, A. J. Heeger, and F. Wudl, *Chem. Rev.* **88**, 183 (1988).
12. J. R. Reynolds, *Chemtech* **18**, 440 (1988).
13. J. R. Reynolds and M. Pomerantz, in T. A. Skotheim, ed., *Electroresponsive Molecular and Polymeric Materials*, Marcel Dekker, New York, 1991.
14. W. M. Wright and G. W. Woodham, in J. M. Margolis, ed., *Conductive Polymers and Plastics*, Chapman and Hall, New York, 1989.
15. A. J. Polak, in Ref. 14.
16. T. Ito, H. Shirakawa, and S. Ikeda, *J. Polym. Sci., Polym. Chem. Ed.* **12**, 11 (1974).
17. N. Theophilou and co-workers, *Eur. Polym. J.* **23**, 11 (1987).
18. *Ibid.*, p. 15.
19. I. Kminek and I. Trekoval, *Makromol. Chem. Rapid Commun.* **7**, 53 (1986).
20. H. Naarman, *Synth. Met.* **17**, 223 (1987).
21. H. Tanaka and T. Danno, *Synth. Met.* **17**, 545 (1987).
22. H. Naarman and N. Theophilou, *Synth. Met.* **22**, 1 (1987).
23. G. Lugli, U. Pedretti, and G. Perego, *J. Polym. Sci. Polym. Lett. Ed.* **23**, 129 (1985).

24. D. M. Ivory and co-workers, *J. Chem. Phys.* **71**, 1506 (1979).

25. A. F. Diaz, K. K. Kanazawa, and G. P. Gardini, *J. Chem. Soc., Chem. Commun.*, 635 (1979).

26. J. H. Edwards and W. J. Feast, *Polymer* **21**, 595 (1980).

27. J. H. Edwards, W. J. Feast, and D. C. Bott, *Polymer* **25**, 395 (1984).

28. G. Lieser, G. Wegner, R. Weizenhofer, and L. Brombacher, *Polym. Prepr., Am. Chem. Soc. Div. Polym. Chem.* **25**, 221 (1984).

29. D. White and D. C. Bott, *Polym. Commun.* **25**, 98 (1984).

30. W. J. Feast and J. N. Winter, *J. Chem. Soc., Chem. Commun.*, 202 (1985).

31. T. M. Swager, D. A. Dougherty, and R. H. Grubbs, *J. Am. Chem. Soc.* **110**, 2973 (1988).

32. T. M. Swager and R. H. Grubbs, *J. Am. Chem. Soc.* **111**, 4413 (1989).

33. J. C. W. Chien, G. E. Wnek, F. E. Karasz, and J. A. Hirsch, *Macromolecules* **14**, 479 (1981).

34. G. F. Dandreaux, M. E. Galvin, and G. E. Wnek, *J. Phys. (Les Ulis, Fr.)* **C3**, 135 (1983).

35. M. E. Galvin and G. E. Wnek, *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **24**(2), 14 (1983).

36. M. E. Galvin and G. E. Wnek, *Polym. Bull.* **13**, 109 (1985).

37. M. Aldissi and A. R. Bishop, *Polymer* **26**, 622 (1985).

38. J. A. Stowell, A. J. Amass, M. S. Beevers, and T. R. Farren, *Makromol. Chem.* **188**, 1635 (1987).

39. *Ibid.*, p. 2535.

40. G. L. Baker and F. S. Bates, *J. Phys. (Les Ulis, Fr.)* **C3**, 11 (1983).

41. F. S. Bates and G. L. Baker, *Macromolecules* **16**, 704 (1983).

42. G. L. Baker and F. S. Bates, *Macromolecules* **17**, 2619 (1984).

43. S. Destri, M. Catellani, and A. Bolognesi, *Makromol. Chem., Rapid Commun.* **5**, 353 (1984).

44. P. Kovacic and C. Wu, *J. Polym. Sci.* **47**, 448 (1960).

45. P. Kovacic and M. B. Jones, *Chem. Rev.* **87**, 357 (1987).

46. C. E. Brown and co-workers, *J. Polym. Sci., Polym. Chem. Ed.* **24**, 255 (1986).

47. T. Yamamoto, Y. Hayashi, and Y. Yamamoto, *Bull. Chem. Soc. Jpn.* **51**, 2091 (1978).

48. D. G. H. Ballard, A. Courtis, I. M. Shirley, and S. C. Taylor, *J. Chem. Soc., Chem. Commun.*, 954 (1983).

49. D. G. H. Ballard, A. Courtis, I. M. Shirley, and S. C. Taylor, *Macromolecules* **21**, 294 (1988).

50. P. R. McKean and J. K. Stille, *Macromolecules* **20**, 1787 (1987).

51. M. Rehahn, A.-D. Schlöter, G. Wegner, and W. J. Feast, *Polymer* **30**, 1054 (1989).

52. *Ibid.*, p. 1060.

53. M. Fukuda, K. Sawada, and K. Yoshino, *Jpn. J. Appl. Phys.* **28**, L1433 (1989).

54. A. Dall'Olio, Y. Dascola, V. Varacco, and V. Bocchi, *C. R. Acad. Sci., Ser. C* **267**, 433 (1968).

55. M. Salmon, A. F. Diaz, J. A. Logan, M. Krounbi, and J. Bargon, *Mol. Cryst. Liq. Cryst.* **83**, 265 (1982).

56. A. F. Diaz, A. Martinez, K. K. Kanazawa, and M. Salmon, *J. Electroanal. Chem.* **130**, 181 (1981).

57. T. C. Clarke, J. C. Scott, and G. B. Street, *IBM J. Res. Dev.* **27**, 313 (1983).

58. J. C. Scott, P. Pflüger, M. T. Krounbi, and G. B. Street, *Phys. Rev. B* **28**, 2140 (1983).

59. G. B. Street and co-workers, *J. Phys. (Paris)* **C3**, 599 (1983).

60. P. Pflüger and G. B. Street, *J. Phys. (Paris)* **C3**, 609 (1983).

61. A. F. Diaz and B. Hall, *IBM J. Res. Dev.* **27**, 342 (1983).

62. K. J. Wynne and G. B. Street, *Macromolecules* **18**, 2361 (1985).

63. N. Bates, M. Cross, R. Lines, and D. Walton, *J. Chem. Soc., Chem. Commun.*, 871 (1985).

64. L. F. Warren and D. P. Anderson, *J. Electrochem. Soc.* **134**, 101 (1987).

65. D. T. Glatzhofer, J. Ulanski, and G. Wegner, *Polymer* **28**, 229 (1987).

66. T. Shimidzu, A. Ohtani, T. Iyoda, and K. Honda, *J. Electroanal. Chem.* **224**, 123 (1987).

67. M. B. Gieselman and J. R. Reynolds, *Macromolecules* **23**, 3118 (1990).

68. M. R. Bryce, A. Chissel, P. Kathirgamanathan, D. Parker, and N. R. M. Smith, *J. Chem. Soc., Chem. Commun.*, 466 (1987).

69. **Eur. Pat. Appl. EP253,595** (Jan. 1988), K. Poopathy and N. R. M. Smith.

70. J. Ruhe, T. Ezquerra, and G. Wegner, *Makromol. Chem., Rapid Commun.* **10**, 103 (1989).

71. H. Masuda, S. Tanaka, and K. Kaeriyama, *J. Chem. Soc., Chem. Commun.*, 725 (1989).

72. A. Deronzier and J.-C. Moutet, *Acc. Chem. Res.* **22**, 249 (1989).

73. A. F. Diaz and co-workers, *J. Electroanal. Chem.* **133**, 233 (1982).

74. G. Bidan and M. Guglielmi, *Synth. Met.* **15**, 49 (1986).

75. J. R. Reynolds, P. A. Porapatic, and R. L. Toyooka, *Macromolecules* **20**, 958 (1987).

76. A. O. Patil, Y. Ikenoue, F. Wudl, and A. J. Heeger, *J. Am. Chem. Soc.* **109**, 1858 (1987).

77. A. O. Patil and co-workers, *Synth. Met.* **20**, 151 (1987).

78. N. S. Sundaresan, S. Basak, M. Pomerantz, and J. R. Reynolds, *J. Chem. Soc., Chem. Commun.*, 621 (1987).

79. P. G. Pickup, *J. Electroanal. Chem.* **225**, 273 (1987).

80. J. R. Reynolds, N. S. Sundaresan, M. Pomerantz, S. Basak, and C. K. Baker, *J. Electroanal. Chem.* **250**, 355 (1988).

81. M. Sato, S. Tanaka, and K. Kaeriyama, *J. Chem. Soc., Chem. Commun.*, 873 (1986).

82. R. L. Elsenbaumer, K.-Y. Jen, and R. Oboodi, *Synth. Met.* **15**, 169 (1986).

83. K.-Y. Jen, G. G. Miller, and R. L. Elsenbaumer, *J. Chem. Soc., Chem. Commun.*, 1346 (1986).

84. R. Sugimoto, S. Takeda, H. B. Gu, and K. Yoshino, *Chem. Express* **1**, 635 (1986).

85. M. Lemaire, R. Garreau, F. Garnier, and J. Roncali, *New J. Chem.* **11**, 703 (1987).

86. K. Kaeriyama, M. Sato, and S. Tanaka, *Synth. Met.* **18**, 233 (1987).

87. S. Hotta, *Synth. Met.* **22**, 103 (1987).

88. M. Sato, S. Tanaka, K. Kaeriyama, and F. Tomanaga, *Polymer* **28**, 107 (1987).

89. J. Roncali, R. Garreau, D. Delabouglise, F. Garnier, and M. Lemaire, *Synth. Met.* **28**, C341 (1989).

90. A. O. Patil, *Synth. Met.* **28**, C495 (1989).

91. M. Lemaire, D. Delabouglise, R. Garreau, A. Guy, and J. Roncali, *J. Chem. Soc., Chem. Commun.*, 658 (1988).

92. D. Kotkar, V. Joshi, and P. H. Ghosh, *J. Chem. Soc., Chem. Commun.*, 917 (1988).

93. R. L. Blakespoor and L. L. Miller, *J. Chem. Soc., Chem. Commun.*, 90 (1985).

94. S. Tanaka, M. Sato, and K. Kaeriyama, *Polym. Commun.* **26**, 303 (1985).

95. K. Kaeriyama, S. Tanaka, M. Sato, and K. Hamada, *Synth. Met.* **28**, C611 (1989).

96. R. L. Elsenbaumer, K.-Y. Jen, G. G. Miller, and L. W. Shacklette, *Synth. Met.* **18**, 277 (1987).

97. J. P. Ruiz, K. Nayak, D. S. Marynick, and J. R. Reynolds, *Macromolecules* **22**, 1231 (1989).

98. J. R. Reynolds and co-workers, *Synth. Met.* **28**, C621 (1989).

99. J. D. Capistran, D. R. Gagnon, S. Antoun, R. W. Lenz, and F. E. Karasz, *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **25**, 282 (1984).

100. D. R. Gagnon, J. D. Capistran, F. E. Karasz, and R. W. Lenz, *Polym. Bull.* **12**, 293 (1984).

101. I. Murase, T. Ohnishi, T. Noguchi, and M. Hirooka, *Polym. Commun.* **25**, 327 (1984).

102. I. Murase, T. Ohnishi, T. Noguchi, M. Hirooka, and S. Murakami, *Mol. Cryst. Liq. Cryst.* **118**, 333 (1985).

103. M. Kanbe and M. Okawara, *J. Polym. Sci. Part A-1* **6**, 1058 (1968).

104. **U.S. Pat. 3,401,152** (1968), R. A. Wessling and R. G. Zimmerman.

105. **U.S. Pat. 3,706,677** (1972), R. A. Wessling and R. G. Zimmerman.

106. R. W. Lenz, C.-C. Han, J. Stenger-Smith, and F. E. Karasz, *J. Polym. Sci., Polym. Chem. Ed.* **26**, 3241 (1988).

107. I. Murase, T. Ohnishi, T. Noguchi, and M. Hirooka, *Synth. Met.* **17**, 639 (1987).

108. S. Antoun, D. R. Gagnon, F. E. Karasz, and R. W. Lenz, *Polym. Bull.* **15**, 181 (1986).

109. K.-Y. Jen, L. W. Shacklette, and R. L. Elsenbaumer, *Synth. Met.* **22**, 179 (1987).

110. C.-C. Han, and R. L. Elsenbaumer, *Synth. Met.* **30**, 123 (1989).

111. S. H. Askari, S. D. Rughooputh, and F. Wudl, *Synth. Met.* **29**, E129 (1989).

112. S. H. Askari, S. D. Rughooputh, and F. Wudl, *Polym. Mater. Sci. Eng.* **59**, 1068 (1988).

113. K.-Y. Jen, M. Maxfield, L. W. Shacklette, and R. L. Elsenbaumer, *J. Chem. Soc., Chem. Commun.*, 309 (1987).

114. R. L. Elsenbaumer and co-workers, in J. Kuzmany, M. Mehring, and S. Roth, eds., *Electronic Properties of Conjugated Polymers*, Springer-Verlag, Berlin, 1987, p. 400.

115. **PCT Int. Pat. WO 88 00,954** (1988), K.-Y. Jen, R. L. Elsenbaumer, and L. W. Shacklette.

116. I. Murase, T. Ohnishi, T. Noguchi, and M. Hirooka, *Polym. Commun.* **28**, 229 (1987).

117. **Ger. Pat. DE 3,704,411** (1987), I. Murase, T. Ohnishi, and T. Noguchi.

118. S. Yamada, S. Tokito, T. Tsutsui, and S. Saito, *J. Chem. Soc., Chem. Commun.*, 1448 (1987).

119. K.-Y. Jen, T. R. Jow, and R. L. Elsenbaumer, *J. Chem. Soc., Chem. Commun.*, 1113 (1987).

120. K.-Y. Jen and co-workers, *Mol. Cryst. Liq. Cryst.* **160**, 69 (1988).

121. K.-Y. Jen, H. Eckhardt, T. R. Jow, L. W. Shacklette, and R. L. Elsenbaumer, *J. Chem. Soc., Chem. Commun.*, 215 (1988).

122. J.-I. Jin, H.-K. Shim, and R. W. Lenz, *Synth. Met.* **29**, E53 (1989).

123. H.-K. Shim, R. W. Lenz, and J.-I. Jin, *Makromol. Chem.* **190**, 389 (1989).

124. F.-L. Lu, F. Wudl, M. Nowak, and A. J. Heeger, *J. Am. Chem. Soc.* **108**, 8311 (1986).

125. L. W. Shacklette, J. F. Wolf, S. Gould, and R. H. Baughman, *J. Chem. Phys.* **88**, 3955 (1988).

126. F. Wudl and co-workers, *J. Am. Chem. Soc.* **109**, 3677 (1987).

127. A. G. MacDiamid, J.-C. Chiang, A. E. Richter, N. L. D. Somasiri, and A. J. Epstein, in L. Alcacer, ed., *Conducting Polymers*, Reidel, Dordrecht, Holland, 1987, p. 105.

128. E. M. Genies, M. Lapkowski, and J. F. Penneau, *J. Electroanal. Chem.* **249**, 97 (1988).

129. Y. Wei, W. W. Focke, G. E. Wnek, A. Ray, and A. G. MacDiamid, *J. Phys. Chem.* **93**, 495 (1989).

130. J. Guay, M. Leclerc, and L. H. Dao, *J. Electroanal. Chem.* **251**, 31 (1988).

131. *Ibid.*, p. 21.

132. M. Leclerc, J. Guay, and L. H. Dao, *Macromolecules* **22**, 641 (1989).

133. L. H. Dao, J. Guay, M. Leclerc, and J.-W. Chevalier, *Synth. Met.* **29**, E377 (1989).

134. L. H. Dao, J. Guay, and M. Leclerc, *Synth. Met.* **29**, E383 (1989).

135. J.-W. Chevalier, J.-Y. Bergeron, and L. H. Dao, *Polym. Commun.* **30**, 308 (1989).

136. M. E. Galvin and G. E. Wnek, *Polym. Commun.* **23**, 795 (1982).

137. M. E. Galvin and G. E. Wnek, *J. Polym. Sci., Polym. Chem. Ed.* **21**, 2727 (1983).

138. M. F. Rubner, S. K. Tripathy, J. Georger, and P. Cholewa, *Macromolecules* **16**, 870 (1983).

139. J. M. Machado, F. E. Karasz, and R. W. Lenz, *Polymer* **29**, 1412 (1988).

140. J. B. Schlenoff, J. M. Machado, P. J. Glatkowski, and F. E. Karasz, *J. Polym. Sci., Polym. Phys. Ed.* **26**, 2247 (1988).

141. J. M. Machado, J. B. Schlenoff, and F. E. Karasz, *Macromolecules* **22**, 1964 (1989).

142. S. Hotta, S. D. Rughooputh, and A. J. Heeger, *Synth. Met.* **22**, 79 (1987).

143. M. B. Gieselman and J. R. Reynolds, unpublished data, 1992.

144. J.-E. H sterholm and co-workers, *Synth. Met.* **28**, C435 (1989).

145. H. Isotalo and co-workers, *Synth. Met.* **28**, C461 (1989).

146. J. Laakso, J.-E. H sterholm, and P. Nyholm, *Synth. Met.* **28**, C467 (1989).

147. J.-O. Nilsson and co-workers, *Synth. Met.* **28**, C445 (1989).

148. S. E. Lindsey and G. B. Street, *Synth. Met.* **10**, 67 (1984).

149. M.-A. DePaoli, R. J. Waltman, A. F. Diaz, and J. Bargon, *J. Chem. Soc., Chem. Commun.*, 1015 (1984).

150. O. Niwa and T. Tamamura, *J. Chem. Soc., Chem. Commun.*, 817 (1984).

151. M.-A. DePaoli, R. J. Waltman, A. J. Diaz, and J. Bargon, *J. Polym. Sci., Polym. Chem. Ed.* **23**, 1687 (1985).

152. O. Niwa, M. Kakuchi, and T. Tamamura, *Polym. J.* **19**, 1293 (1987).

153. O. Niwa, M. Hikita, and T. Tamamura, *Appl. Phys. Lett.* **46**, 444 (1985).

154. F.-R. F. Fan and A. J. Bard, *J. Electrochem. Soc.* **133**, 301 (1986).

155. R. M. Penner and C. R. Martin, *J. Electrochem. Soc.* **133**, 310 (1986).

156. H. Yoneyama, T. Hirai, S. Kuwabata, and O. Ikeda, *Chem. Lett.*, 1243 (1986).

157. G. Nagasubramanian, S. Di Stefano, and J. Mascanin, *J. Phys. Chem.* **90**, 447 (1986).

158. R. M. Penner and C. R. Martin, *J. Electrochem. Soc.* **132**, 514 (1985).

159. T. Shimidzu, A. Ohtani, T. Iyoda, and K. Honda, *J. Chem. Soc., Chem. Commun.*, 1415 (1986).

160. *Ibid.*, 1987, p. 327.

161. R. E. Nofle and D. Pletcher, *J. Electroanal. Chem.* **227**, 229 (1987).

162. L. L. Miller and Q.-X. Zhou, *Macromolecules* **20**, 1594 (1987).

163. T. Shimidzu, A. Ohtani, and K. Honda, *J. Electroanal. Chem.* **251**, 323 (1988).

164. A. Ohtani, M. Abe, H. Higuchi, and T. Shimidzu, *J. Chem. Soc., Chem. Commun.*, 1545 (1988).

165. K. Honda, Y. Kishigami, and K. Takemoto, *Makromol. Chem., Rapid Commun.* **9**, 101 (1988).

166. S. J. Jasne and C. K. Chiklis, *Synth. Met.* **15**, 175 (1986).

167. J. Edwards, R. Fisher, and B. Vincent, *Makromol. Chem., Rapid Commun.* **4**, 393 (1983).

168. S. P. Armes and B. Vincent, *J. Chem. Soc., Chem. Commun.*, 288 (1987).

169. S. P. Armes, J. F. Miller, and B. Vincent, *J. Colloid Interfac. Sci.* **118**, 410 (1987).

170. R. J. Cohen and A. J. Glick, *Phys. Rev. Lett. B* **42**(12), 7659 (1990).

171. J. L. Bredas, J. C. Scott, K. Yakushi, and G. B. Street, *Phys. Rev. Lett. B* **30**, 1023 (1982).

172. S. Ikehata and co-workers, *Phys. Rev. Lett.* **45**, 423 (1980).

173. J. C. Scott, M. Krounbi, P. Pfluger, and G. B. Street, *Phys. Rev. Lett. B* **28**, 2140 (1983).

174. M. Peo and co-workers, *Solid State Commun.* **35**, 119 (1980).

175. G. Zotti and G. Schiavon, *Synth. Met.* **31**, 347 (1989).

176. G. Froyer and co-workers, *Mol. Cryst. Liq. Cryst.* **118**, 267–272 (1985).

177. G. Zotti and G. Schiavon, *Chem. Mater.* **3**, 62–65 (1991).

178. R. H. Baughman, J. L. Bredas, R. R. Chance, R. L. Elsenbaumer, and L. W. Shacklette, *Chem. Rev.* **82**, 209 (1982).
179. A. R. Bishop, D. K. Campbell, and K. Fesser, *Mol. Cryst. Liq. Cryst.* **77**, 253 (1981).
180. W. P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698 (1979).
181. J. R. Reynolds, J. P. Ruiz, A. D. Child, K. Nayak, and D. S. Marynick, *Macromolecules* **24**, 678–687 (1991).
182. G. Touilllon and F. Garnier, *J. Electroanal. Chem.* **135**, 173–178 (1982).
183. K. J. Wynne and G. B. Street, *Macromolecules* **18**, 2361 (1985).
184. C. K. Baker, Y. J. Qiu, and J. R. Reynolds, *J. Phys. Chem.* **95**, 4446 (1991).
185. J. R. Reynolds, M. B. Gieselman, Y. J. Qiu, and M. Pyo, *Proc. Am. Chem. Soc., Div. Polym. Sci. Eng.* **64**, 202 (1991).
186. T. C. Chung, J. H. Kaufman, A. J. Heeger, and F. Wudl, *Phys. Rev. B* **30**, 702 (1984).
187. S. N. Hoier, D. D. Ginley, and S.-M. Park, *J. Electrochem. Soc.* **135**, 91–94 (1988).
188. S. Wang, T. Kawai, K. Yoshino, K. Tanaka, and T. Yamabe, *Jpn. J. Appl. Phys.* **29**, L2264–2267 (1990).
189. K. Nishioka, S. Wang, and K. Yoshino, *Synth. Met.* **41–43**, 815–818 (1991).
190. K. F. Voss and co-workers, *Phys. Rev. B* **43**(6), 5109 (1991).
191. H. Eckhart, L. W. Shacklette, K.-Y. Jen, and R. L. Elsenbaumer, *J. Chem. Phys.* **91**, 1303 (1989).
192. F. Wudl, M. Kobayashi, and A. J. Heeger, *J. Org. Chem.* **49**, 3382 (1984).
193. J. L. Bredas, J. C. Scott, K. Yakushi, and G. B. Steet, *Phys. Rev. B* **30**, 1023 (1984).
194. A.-C. Chang, R. L. Blankespoor, and L. L. Miller, *J. Electroanal. Chem.* **236**, 239 (1987).
195. G. Daoust and M. Leclerc, *Macromolecules* **24**, 455 (1991).
196. N. Basescu and co-workers, *Nature* **327**(4), 403 (1987).
197. Y. Wang and M. F. Rubner, *Synth. Met.* **39**, 153 (1990).
198. Y. Wang, M. F. Rubner, and L. J. Buckley, *Synth. Met.* **41–43**, 1103 (1991).
199. G. Gustafsson, O. Inganas, J. O. Nilsson, and B. Liedberg, *Synth. Met.* **26**, 297 (1988).
200. U. G. Wettermark, G. A. Worrell, C. S. Chen, M. B. Gieselman, and J. R. Reynolds, *Proc. Am. Chem. Soc., Div. Polym. Mats. Sci. Eng.* **64**, 267 (1991).
201. M. Angelopoulos, J. M. Shaw, R. D. Kaplan, and S. Perreault, *J. Vac. Sci. Tech. B* **7**, 1519 (1989).
202. M. Angelopoulos and J. M. Shaw, *Microelectronic Eng.* **13**, 515 (1991).
203. W. S. Huang, M. Angelopoulos, J. R. White, and J. M. Park, *Mol. Cryst., Liq. Cryst.* **189**, 227 (1990).
204. P. J. Nigrey, D. Macinnes, Jr., D. P. Nairns, A. G. MacDiarmid, and A. J. Heeger, *J. Electrochem. Soc.* **128**(8), 1651 (1981).
205. H. Munstedt and co-workers, presented at the *International Conference on the Science and Technology of Synthetic Metals*, Sante Fe, N.M., 1988.
206. L. W. Shacklette, T. R. Jow, M. Maxfield, and R. Hatami, *Synth. Met.* **28**, C655 (1989).
207. T. Matsunaga, H. Daifuku, T. Nakajima, and T. Kawagoe, *Polymers for Adv. Technol.* **1**, 33 (1990).
208. R. V. Gregory, W. C. Kimbell, and H. H. Kuhn, *Synth. Met.* **28**, C823 (1989).
209. R. V. Gregory, W. C. Kimbell, and H. H. Kuhn, *J. Coated Fabrics* **20**, 1 (1991).
210. R. H. Baughman, R. L. Elsenbaumer, Z. Iqbal, G. G. Miller, and H. Eckhardt, *Proceedings of the International Winter School on Conducting Polymers*, Kirchberg, Austria, Mar. 1987.
211. G. Gustafsson and I. Lundstrom, *Proceedings of the Second International Meeting on Chemical Sensors*, Bordeaux, France, July 1986.
212. N. C. Foulds and C. R. Lowe, *J. Chem. Soc., Faraday Trans. 1* **82**, 1259 (1986).
213. N. F. Colaneri and L. W. Shacklette, *IEEE Trans. Instr. Meas.* **41**, 291 (1992).
214. W. Liang and C. R. Martin, *Chem. Mater.* **3**, 390 (1991).
215. M. R. Anderson, B. R. Mattes, H. Reiss, and R. B. Kaner, *Science* **252**, 1412 (1991).
216. J. H. Burroughes and co-workers, *Nature* **347**, 539 (1990).
217. D. Braun, A. J. Heeger, and H. Kroemer, *J. Electron. Mater.* **20**, 945 (1991).
218. Y. Ohmori, M. Uchida, K. Muro, and K. Yoshino, *Jpn. J. Appl. Phys.* **30**, L1938 (1991).

John R. Reynolds
Andrew D. Child
University of Florida

Melinda B. Gieselman
3M Specialty Adhesives and Chemicals

ELECTROANALYTICAL TECHNIQUES

Electroanalysis employs electrochemical cells. The most common electrochemical cells are batteries (qv) which consist of two electrodes: an anode and a cathode. The electrochemical principles governing batteries and electroanalytically useful cells are the same. The typical electroanalytical cell has at least two electrodes, a working electrode and a reference electrode. The working electrode may serve as either anode or cathode depending on applied voltage, or as a source or sink for ion exchange (qv) as in the case of ion-selective electrodes. The reference electrode invariably serves as a source or sink for ions at its interface with the solution, and as a source or sink for electrons at its interface with an external circuit. The solution or electrolyte in batteries has a very high ionic strength in order to carry enormous amounts of current. In electroanalytical cells the solution is the sample and must typically contain supporting electrolyte, in the form of electrochemically inert salts, in order to support much lower current densities. Supporting electrolytes may also be used to define ionic strength.

Cells useful for electroanalysis typically consist of two or more electrodes dipping into the sample, ie, the solution to be analyzed. Sample solutions can range from water to blood but can also include virtually all organic solvents. Analyte concentrations are usually in the picomolar to millimolar range. Numerous choices exist for working electrodes, ranging from electron exchangers to ion exchangers. These may be metals of many kinds, eg, Pt, Au, Ag, and Hg; semiconductors (qv), eg, Si, Ge, and TiO_2 ; or plastics, ion exchangers and sequesterers in poly(vinyl chloride) (PVC). Choices for reference electrodes are more limited. The problems involving reference electrodes are often more profound, and availability of the right reference electrode may ultimately dictate the feasibility of an assay.

Reference electrodes are inherently unstable. These electrodes drift, leak, become foul or plugged, and frequently need to be replaced.

Electrochemical cells may be used in either active or passive modes, depending on whether or not a signal, typically a current or voltage, must be actively applied to the cell in order to evoke an analytically useful response. Electroanalytical techniques have also been divided into two broad categories, static and dynamic, depending on whether or not current flows in the external circuit (1). In the static case, the system is assumed to be at equilibrium. The term dynamic indicates that the system has been disturbed and is not at equilibrium when the measurement is made. These definitions are often inappropriate because active measurements can be made that hardly disturb the system and passive measurements can be made on systems that are far from equilibrium. The terms static and dynamic also imply some sort of artificial time constraints on the measurement. Active and passive are terms that nonelectrochemists seem to understand more readily than static and dynamic.

Active Techniques

Active techniques are classified by the method of collection and display of data. If a voltage is applied to the cell and the resultant current measured and displayed as a function of time, the technique is called chronoamperometry (2). "Chrono" is a combining form from the Greek *chronos*, meaning "time". If a current is applied to the cell and the resultant voltage is measured and displayed as a function of time, the technique is called chronopotentiometry (2). Similarly, if current is measured and displayed as a function of applied potential, the technique is called voltammetry (3). Subcategorizing depends on such things as the geometry and size of the working electrode, its physical treatment, usually its rotation, and the functionality of the applied waveform. Therefore there is rotating disk voltammetry (3), cyclic voltammetry (4), pulse voltammetry (1), etc. Even post-collection treatment of the data may result in a renaming of the technique. Integration of the current, either during analogue data acquisition or digitally from computer stored currents, results in chronocoulometry (2). The functionality of the time response of the charge, measured in coulombs, provides the concentration of the electroactive species. If the electrode is small, eg, 0.01 cm^2 or so, the number of moles of the sample electrolyzed is insignificant relative to its total number of moles. Extremely large ($>1\text{ cm}^2$) electrodes may be used to completely electrolyze the solution, in which case the technique may be termed coulometry (5).

The question of what happens when an electrical signal is applied to an electrochemical cell needs to be answered with respect to the three components of the cell: the working electrode, the reference electrode, and the sample itself.

The Working Electrode. If the applied signal is a voltage and the intent is to measure a flow of current, the outcome depends on the sign and magnitude of the applied voltage and on the chemistries of the solution and the electrode. If the electrode is a metal that appears high in the electromotive series of metals, eg, copper (qv), mercury (qv), or silver (see SILVER AND SILVER ALLOYS), and the voltage is positive with respect to the solution, then the metal may dissolve. On the other hand, mercury is an extremely useful metal for electroanalysis because at negative potentials metal ions in solution, such as Cu^+ , Cu^{2+} , Cd^{2+} , Fe^{2+} , Fe^{3+} , Mn^{2+} , and Ni^{2+} may be reduced to the corresponding metals, which then dissolve in the mercury. If the dropping mercury electrode (DME) is used as the working electrode to replace the amalgam, and the current that flows during the reduction is monitored as a function of applied potential, then the technique is a form of voltammetry called polarography (6). If only one drop is used, ie, a hanging mercury drop electrode (HME), then the metal ions of interest may be reduced at a suitably negative potential for a lengthy period (minutes) of time and then the potential may be linearly scanned in a positive direction to reoxidize the amalgam. This results in a series of current peaks appearing at different voltages on the voltammogram, each corresponding to a different species of metal ion in the solution. The maximum current of each peak is then indicative of the concentration of the corresponding ion. This technique is called anodic stripping voltammetry (7). Anodic means that the electrochemical process occurring at the surface of the electrode is oxidation. If reduction occurs, the electrode process is cathodic.

Polarography and related techniques are not confined to analysis of metal ions. Organic redox reactions can also be utilized for polarographic electroanalysis (8,9). The use of mercury has the advantage of high reproducibility. Each drop presents a new electrode surface to the solution, and anything that may have happened to the previous drop, such as the adsorption of product and concomitant fouling of the droplet surface, becomes inconsequential. The disadvantage of mercury usage lies in the relative ease with which mercury can be oxidized. Mercury has a small anodic limit, ie, the maximum positive voltage that can be applied without causing oxidation, and other electrodes having more positive anodic limits are often more appropriate for use in organic solvents such as acetonitrile, dimethylsulfoxide, methylene chloride, etc. More noble metals such as gold (see GOLD AND GOLD COMPOUNDS), and especially platinum (see PLATINUM GROUP METALS), are useful for these purposes. The anodic limit in these cases is more often defined by oxidation of the solvent, not the electrode.

Charging Current. In most cases, application of a voltage to an electrode is intended to produce an analytically useful current that depends solely on the concentration of the analyte. Unfortunately, current flows even in the complete absence of the analyte. Thus, the current may have nothing to do with the electroactive species in the sample. This charging current must be circumvented or otherwise compensated.

The applied voltage injects charge (electrons) onto the electrode that must be balanced by charge (ions) in solution. The ions, present in the solution as supporting electrolyte, cannot approach the surface of the electrode any closer than their radii permit. This plane of closest approach, also called the inner Helmholtz plane (IHP), defines one plate of a capacitor (Fig. 1). The electrode defines the other plate and solvent molecules adsorbed to the surface (most of the electrode's surface is covered by adsorbed solvent molecules) provide the dielectric material for the capacitor, which is called the compact double-layer capacitance. An outer Helmholtz plane (OHP), defined by the line of centers of solvated ions, may also contribute to the compact double-layer capacitance. There is also a diffuse double-layer capacitance, which generally becomes important only at low ionic strength. This latter capacitance results from nonelectroneutral concentrations of anions and cations in solution that respond to the charge on the electrode, one being attracted and the other repelled. Additionally, there are diffusive forces. The higher concentration ion is trying to diffuse from the electrode region back out into the solution; the lower concentration ion is trying to diffuse in the opposite direction. The equilibrium between electrostatic and diffusive forces establishes the

diffuse double layer and associated capacitance. At the ionic strengths necessary for most electroanalyses the diffusive forces are overwhelmed by the electrostatic force. Thus the total double-layer capacitance, which is the sum of the two contributions, added together as capacitors in series, equals the compact double-layer capacitance.

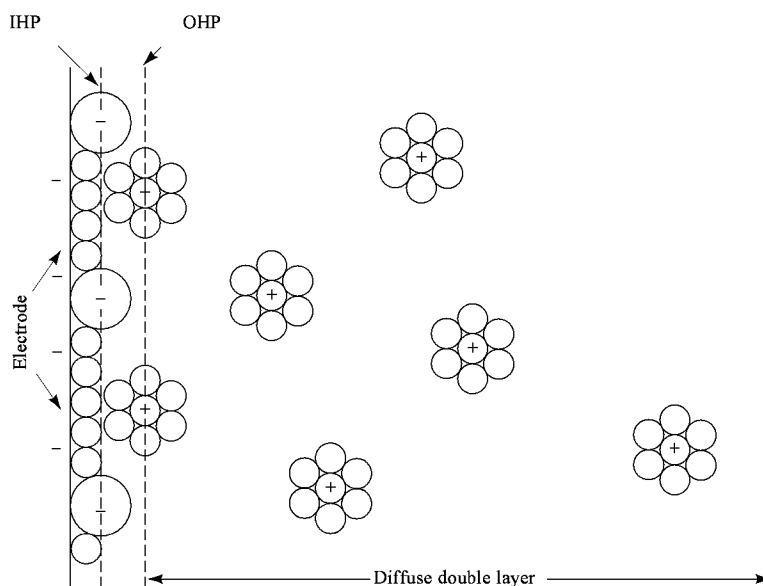


Fig. 1. Schematic representation of the electrochemical or diffuse double layer showing the inner (IHP) and outer (OHP) Helmholtz planes and the electrode. The uncharged circles represent solvent molecules.

From an electroanalytical point of view, the double-layer capacitance is a nuisance resulting in the charging current, which has no analytical value. The charging current must decay to zero, or be otherwise accounted for, before the analytically useful Faradaic current that results from electron exchange with electroactive species in the sample can be measured and used for calibration. Because the double layer is charged by ions that must move through the solution, the time constant for charging is the product of its capacitance (a few μF) and the resistance of the solution (typically $100\ \Omega$). Charging can therefore be very fast. This is an important consideration because there is a relationship between the speed with which a current can be measured and the precision and detection limit of the assay. Various pulse and square-wave techniques, eg, pulse voltammetry and square-wave voltammetry (10), are used to increase the rate of charging and therefore the precision, accuracy, and or speed of the assay.

Faradaic Current. The double layer is a leaky capacitor because Faradaic current flows around it. This leaky nature can be represented by a voltage-dependent resistance placed in parallel and called the charge-transfer resistance. Basically, the electrochemical reaction at the electrode surface consists of four thermodynamically defined states, two each on either side of a transition state. These are (1) oxidized species beyond the diffuse double layer and n electrons in the electrode and (2) oxidized species within the outer Helmholtz plane and n electrons in the electrode, on one side of the transition state; and (3) reduced species within the outer Helmholtz plane and (4) reduced species beyond the diffuse double layer, on the other.

The thermodynamics coupled with the Arrhenius activation theory suggest that the transition state surmounts a voltage-dependent activation energy barrier. The shape of this barrier is described by a parameter, α , that can have values between 0 and 1. If $\alpha = 0.5$ the barrier is symmetrical, if <0.5 it is skewed toward the reduced species, and if >0.5 it is skewed toward the oxidized species (12). The height of the barrier ultimately dictates the rate of charge transfer, and the value of the heterogeneous charge-transfer rate constant, k^0 . These two parameters, α and k^0 , may be used (at least empirically) to describe the charge-transfer process and the accompanying flow of current. If k^0 is large, the barrier is small and surface concentrations of oxidized and reduced species closely obey the Nernst equation, even though large currents may be flowing and the Nernst equation rigorously applies only to equilibrium conditions.

Smaller values of k^0 necessitate the application of voltages greater than those calculated from the Nernst equation to obtain a corresponding set of surface concentrations of electroactive species. These voltages are called overpotentials and indicate chemically related difficulties with the electrolysis. In other words, electron exchange between the electrode and the electroactive species is impeded by the chemistry of the process itself.

Surface concentration ratios of oxidized to reduced species relate directly to the analytically important current densities. In some cases this problem can be solved by simply applying voltages large enough to accommodate the overpotentials. In other cases they are too large to be accommodated, or larger voltages result in spurious currents, breakdown of the solvent, or dissolution of the electrode. Even in the best of cases, overpotentials compromise the already inherently limited resolution of electroanalysis. Electrochemical reactions having small values of k^0 are said to be irreversible. That is, the electron-transfer process is so inefficient that the Nernst equation is not obeyed during a flow of current. Processes that are irreversible on one time scale, however, can become reversible on a longer time scale.

The Reference Electrode. The response of a reference electrode to an applied voltage depends critically on the magnitude of the applied voltage and the resultant current density, the duration of the application, and the nature of the reference electrode itself. Reference electrodes typically consist of a metal contacting one of its halide salts and a saturated potassium halide solution (of this same halide), separated from the sample solution by a salt bridge containing a high (3 M or more) concentration of an equitransferent salt such as KNO_3 . The silver-silver chloride double junction reference electrode illustrated in Figure 2 is such an electrode. It is called a double junction electrode because there are two frits separating two electrolytic solutions. Another popular electrode is the saturated calomel electrode (SCE) in which the silver is replaced by mercury, which is contacted by platinum for connection to the external circuit, and the silver chloride is replaced by a mixture of mercurous and mercuric chlorides, ie, calomel. The standard electrochemical reference electrode is the standard or normal hydrogen electrode (NHE) which consists of a platinum electrode contacted by a stream of hydrogen gas at 101.3 kPa (1 atm) and an aqueous solution of hydrochloric acid at unit activity.

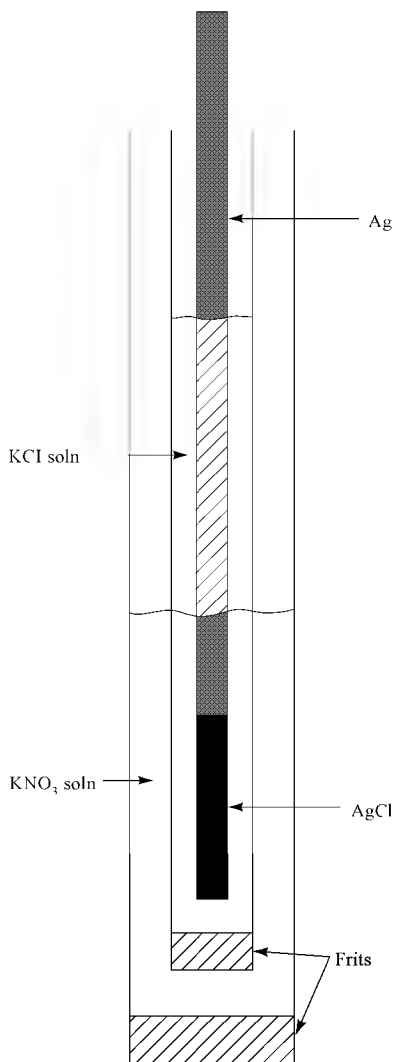


Fig. 2. Schematic of a silver–silver chloride electrode.

There are many variations of reference electrodes. For example the frits may be replaced by fibers, free-flowing capillaries, or Luggin probes (1,12), which are narrow tubes often several centimeters long, through which the salt bridge solution flows. Although the junction must be free-flowing to provide a stable reference potential, the flow rate should be made as small as possible to avoid depletion of the internal filling solutions and or contamination of the sample. Obviously, the reference electrode eventually contaminates a free-standing sample solution and depletes itself of some of its ingredients. This inherent instability precludes many process control (qv) applications for which long-term stability is essential. Instability is not as great for most laboratory applications. Reference electrodes can cause problems even in the absence of an applied potential, but the effects of applied voltages and resultant currents are much less subtle.

The Ag–AgCl electrode (Fig. 2) is a case in point. A positive voltage that results in a prolonged flow of current can eventually convert all of the submerged silver into silver chloride or deplete the concentration of KCl. A negative voltage eventually dissolves the silver chloride coating, which is generally produced by anodization of the silver wire in a chloride solution, and precipitates KCl from its saturated solution. Saturated KCl solutions are not always used, in which case the chloride concentration of this solution may rise noticeably. In any event, the change in composition of the reference electrode results in a change in its reference potential. The instability can be trivial when current densities are low and the current flows for only brief periods of time, or severe when prolonged electrolyses are involved.

The solution to reference electrode instability is the introduction of a third or auxiliary electrode. This particular electrode is intended to carry whatever current is required to keep the potential difference between the working and reference electrodes at a specified value, and virtually all potentiostats (instruments designed specifically for electrochemistry) have this three-electrode configuration. Its use is illustrated in Figure 3.

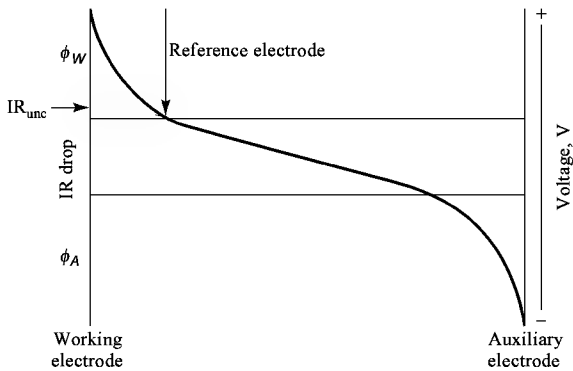


Fig. 3. The three-electrode system. Terms are defined in text.

Whereas the terms voltage and potential have been used almost interchangeably, voltage is the number of volts applied between the auxiliary and working electrodes of the electrochemical cell to maintain a flow of current. Potential, ϕ , is the number of volts (usually millivolts) that drop between any two regions of the cell. ϕ_a is the potential difference between the working and reference electrode. The latter is placed as close to the working electrode as possible in order to minimize the uncompensated IR drop (IR_{unc}) that results from current flow, I , through the solution resistance, R . The IR drop and the

potential drop at the auxiliary electrode are unimportant except for the practical consideration that the potentiostat must be able to apply the correct voltage to the cell to maintain the correct potential at the working electrode.

The three-electrode system serves two important purposes. Because the reference electrode carries no current, but merely measures a potential relative to the working electrode, its stability is not unduly influenced by the electrolysis. Furthermore, because it is placed close to the working electrode the measured potential difference is more nearly representative of the true potential difference between the working electrode and the sample solution. This latter is the significant quantity in electroanalysis.

The resolution of electroanalysis is inherently limited. The entire potential range spanned, from anodic to cathodic limit, cannot be greater than 6 volts and even this range can only be covered using a platinum electrode and noxious non-aqueous solvents like liquid ammonia (qv) and sulfur dioxide. Examination of the Nernst equation reveals that a ± 0.1 V swing about the E^0 of a redox couple results in a 50:1 to 1:50 swing in the ratio of the oxidized to reduced species activities. The total resolution over 6 volts is therefore about $6 \cdot 0.2 \cdot 30$ redox couples or 1 part in 30 resolution. In aqueous solutions the span is usually only about 2 volts, so the resolution is 1 part in 10. Viewed in this context a millivolt is a fairly large quantity, and the precise control of applied potentials, within a few millivolts, may be critical to the success of an electroanalysis. This is especially true if large electrolyses are involved because high current densities result in IR drops that move the applied voltage (Fig. 3) far away from the desired potential difference, ϕ_s . The proper selection and use of the reference electrode, and the importance of the three-electrode system, should therefore not be underestimated.

Whereas electrochemistry has a poor resolution, virtually every analytical technique except chromatography (qv), which has low sensitivity in the form of column or paper chromatography, suffers from lack of resolution when quantifying the analysis of one component in a mixture (see ANALYTICAL METHODS). The resolution of spectroscopic techniques is also limiting and resolving more than two or three components in a complex solution is extremely difficult unless the wavelength range is enormous (see INFRARED AND RAMAN SPECTROSCOPY; SPECTROSCOPY). Atomic absorption spectroscopy, which can resolve hundreds of elements, is an exception. The development and use of hyphenated techniques (see ANALYTICAL METHODS, HYPHENATED INSTRUMENTS), in which the higher resolution of chromatography is married to the sensitivity of another technique, is intended to obtain separation, identification, and quantification of the analyte. An electroanalytical system is liquid chromatography with electrochemical detection (lcec) (1). lcec is extremely versatile, taking advantage of the resolution of chromatography and the high precision of electroanalysis.

The Solution. The responses of working and reference electrodes to applied voltages are important only because this information can be indicative of what goes on in the solution, or at the solution-electrode interface. The distinction between bulk (solution) and interfacial events is basically the distinction between chemical kinetics and charge transfer.

Coulometry. If it can be assumed that kinetic nuances in the solution are unimportant and that destruction of the sample is not a problem, then the simplest action may be to apply a potential to a working electrode having a surface area of several cm^2 and wait until the current decays to zero. The potential should be sufficiently removed from the E^0 of the analyte, ie, about 200 mV, that the electrolysis of an interferent is avoided. The integral under the current vs time curve is a charge equal to $nFCV$, where n is the number of electrons needed to electrolyze the molecule, C is the concentration of the analyte, V is the volume of the solution, and F is the Faraday constant.

Coulometry (5) is not usually the technique employed. Even in the absence of kinetics, the several minutes required for the electrolysis seems excessive and destruction of the sample is not a desirable result. Furthermore, coulometric precision can be exceptionally poor at low concentration, and currents almost never decay to zero because of the trace contaminants present. One has to decide when zero current has been obtained.

Chronoamperometry. In the more general case, a voltage is applied at time zero, the current is ignored for a few microseconds while the charging current decays, and then measured for a few seconds while only that small part of the total concentration adjacent to the surface of a micro working electrode is electrolyzed. For this approach the diameter of the working electrode is typically only a millimeter or so. The currents that flow are only in the microampere range but can nevertheless be easily detected and resolved. According to the Cottrell equation

$$i = \frac{nFC'D^{1/2}}{(\pi t)^{1/2}} \tag{1}$$

plots of current density, i , vs the inverse of the square root of time, t , should be linear having a slope proportional to the concentration of the analyte, C . D is the diffusion coefficient of the analyte. Equation 1 is a solution to Fick's Second Law of Diffusion, having appropriate semi-infinite boundary conditions (1,2,4,12). This technique is chronoamperometry (2).

The chronoamperometric technique illustrates the principle that analytically useful current responses depend critically on the efficiency of analyte mass transport within the solution. The analyte mass transport in turn depends on the efficiency with which an applied voltage can maintain the surface concentrations of oxidized and reduced species at values specified by the Nernst equation. It is generally the case in chronoamperometry that the bulk concentration of one of the species is zero whereas the surface concentration of the other species is forced to zero by the applied potential, but this is not always so.

The situation illustrated in Figure 4 allows both species to coexist. Either of the two sets of curves can be considered the oxidized species; the other is the reduced species. The choice depends on whether oxidation or reduction is occurring at the surface. Assume the upper curve is the reduced species and the lower curve is its oxidized form. An applied voltage has maintained fixed surface concentrations for some period of time including t_1 and t_2 . The concentration profile of the oxidized species decreases at the electrode surface (0 distance) as it is being reduced. Electrolysis therefore results in an increase in the concentration of reduced species at the surface. The concentration profiles approach bulk values far from the surface of the electrode because electrolysis for short times at small electrodes cannot significantly affect the concentrations of species in large volumes of solution.

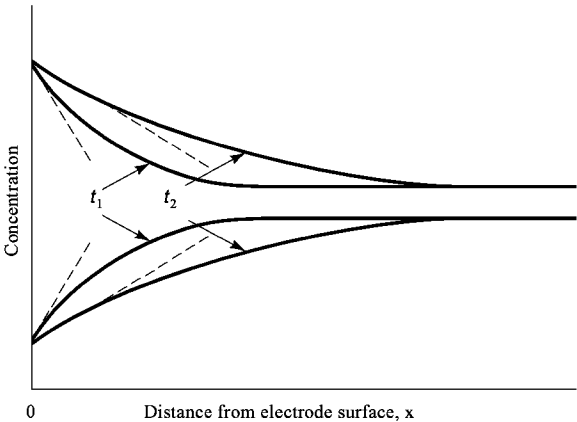


Fig. 4. Concentration profiles during electrolysis of unstirred solutions, where t_1 and t_2 represent two different times.

In this context, the relative terms far, short, small, and large can be defined as follows. Fick's second law of diffusion dictates that the distance, δ , that a species having a diffusion coefficient, D , may diffuse within a period of time, t , is given by (12):

$$\delta = (2Dt)^{1/2}$$

(2)

Aqueous diffusion coefficients are usually on the order of $5 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$. A second is typically a long time to an electrochemist, so $\delta \sim 30 \text{ }\mu\text{m}$. The definition of far is then $30 \text{ }\mu\text{m}$. Short is less than a second, perhaps a few milliseconds. Microseconds are not uncommon. Small, referring to the diameter of the electrode, is about a millimeter for microelectrodes, or perhaps only a few micrometers for ultramicroelectrodes (13).

Small electrodes allow faster measurements and are therefore very popular with electrochemists. Large solution volumes are typically a few tens of milliliters, but electroanalysis can be performed in drops as small as $10 \text{ }\mu\text{L}$ or less because $30 \text{ }\mu\text{m}$ times 1 mm^2 equals only 3% of a $10 \text{ }\mu\text{L}$ drop. Preferred analyte concentration ranges are anywhere from subpicomolar to maybe 10 millimolar, depending on the technique employed.

An important point is that the concentration gradients at the surface, represented by the dashed lines in Figure 4, are proportional to the current densities measured at the applied voltage

$$i = -nFD_o \left. \frac{\partial C_o}{\partial x} \right|_0 = nFD_r \left. \frac{\partial C_r}{\partial x} \right|_0$$

(3)

where C_o , D_o , and C_r , D_r represent the concentration and diffusion constant of the oxidized and reduced species, respectively. These gradients change with the square root of time as diffusion progresses during the time interval that includes t_1 and t_2 . This phenomenon affords an enormous benefit at short times, because the gradients are initially huge. In principle, they are infinite at $t = 0$ when the voltage is first applied, resulting in infinite current flow according to equation 1. In practice, the current is limited by solution resistance and contains the charging current contribution discussed previously. One must then wait the millisecond or less needed to ensure the appropriateness of equation 1 before processing the data. The fact that both profiles are involved, even if the bulk concentration of one of the species is zero, leads to the responses seen in cyclic voltammetry (1,4) and the various kinds of pulse voltammetries (1,10). The currents driven by pulsed voltages are higher and thus more analytically useful.

Voltammetry. Diffusional effects, as embodied in equation 1, can be avoided by simply stirring the solution or rotating the electrode, eg, using the rotating disk electrode (RDE) at high rpm (3,7). The resultant concentration profiles then appear as shown in Figure 5. A time-independent Nernst diffusion layer having a thickness dictated by the laws of hydrodynamics is established. For the RDE,

$$\delta = 1.61 D^{1/3} \omega^{-1/2} \nu^{1/6}$$

(4)

where ω is the radial frequency of rotation and ν is the kinematic viscosity. If $\nu \sim 0.01 \text{ cm}^2 \text{ s}^{-1}$, $\omega = 1000 \text{ s}^{-1}$, and $D = 5 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$, then $\delta \sim 5 \text{ }\mu\text{m}$ corresponding to the diffusion layer thickness calculated from equation 2 where $t \sim 30 \text{ ms}$. However, a steady-state current is measured because stirring maintains the concentration profiles, which may be assumed to be linear within the Nernst diffusion layer. The steady-state currents are therefore directly proportional to the concentration difference between the surface and the bulk, and depend solely on applied potential (ϕ_p) (see Fig. 3). The resultant voltammogram is given by,

$$E = E_{1/2} + \frac{RT}{nF} \ln \left(\frac{i_{l,c} - i}{i - i_{l,a}} \right)$$

(5)

where

$$E_{1/2} = E^{0'} - \frac{RT}{nF} \ln \left(\frac{m_o}{m_r} \right)$$

(6)

cathodic and anodic current limits are given by $i_{l,c}$ and $i_{l,a}$, respectively; R is the ideal gas constant, T is the absolute temperature, and $E^{0'}$ is the formal potential of the redox reaction, ie, the E^0 uncorrected for activity coefficients; m_o and m_r are mass-transfer coefficients (12) for the oxidized and reduced species, respectively, and equal to D/δ . δ depends on the mass-transport properties of the particular technique under consideration (eqs. 2 and 4). For the convective mode of mass transport considered here, for example, $(m_o/m_r) = (D_o/D_r)^{2/3}$, but for the diffusional modes used in polarography and cyclic voltammetry $(m_o/m_r) = (D_o/D_r)^{1/2}$.

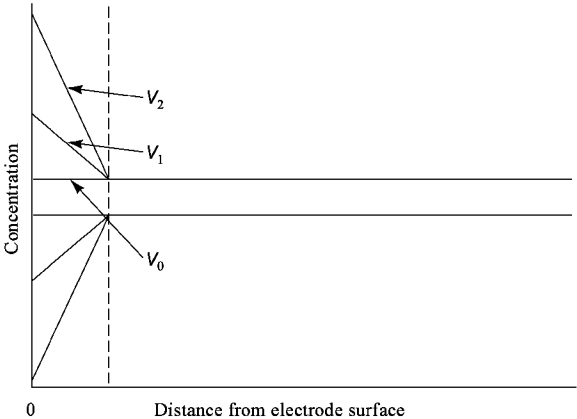


Fig. 5. Concentration profiles at a rotating electrode or in a stirred solution at applied potentials V_0 , V_1 , and V_2 where the dotted line represents the Nernst diffusion layer.

Equation 5 is illustrated schematically in Figure 6. Two additional features to further demonstrate the essential characteristics of this technique are also included. First, a second redox couple is included, having a limiting cathodic current of $i_{l,c,2}$. Second, an irreversible response is included to indicate the overpotential, η , discussed previously. $E_{1/2}$ is the potential corresponding to the average of $i_{l,c}$ and $i_{l,a}$ (see eq. 6) and the potential corresponding to zero current flow is the equilibrium potential. Potential increases negatively (cathodically) to the right in the diagram, so cathodic (reductive) currents are positive. Figure 6 also applies to polarograms, measured at the DME, except that the dropping mercury modulates the current, resulting in a pattern characteristic of polarography (6). The mercury drops essentially stir the solution. Polarography is also closely related to chronoamperometry at an expanding plane (the mercury drop) for a linearly varying voltage.

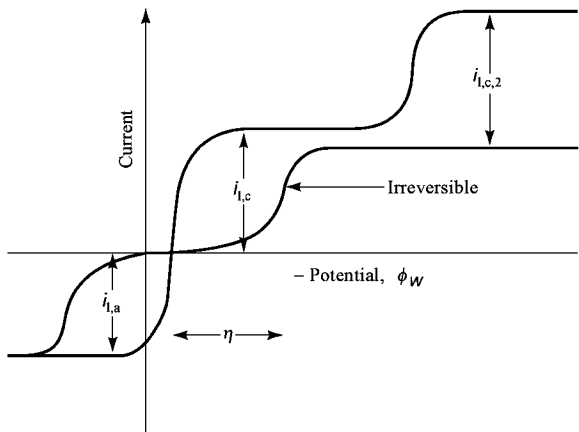


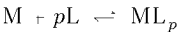
Fig. 6. Voltammogram showing both the reversible and irreversible regions. Terms are defined in text.

The difference between $i_{1,c}$ and $i_{1,a}$ in Figure 6 is a measure of the total concentration of the first redox couple. $i_{1,c,2}$ is a measure of the total concentration of the second redox couple. The equilibrium potential provides the two ratios of oxidized to reduced species according to the Nernst equation. From all this information the concentrations of both oxidized and reduced species can be determined. In this case the E^0 's of the two species are so far apart that their voltammograms are well resolved. The second redox couple must be wholly in its oxidized form whereas the first must be in a ratio equal to the ratio of the absolute values of its limiting currents, corrected by the mass-transfer coefficients.

The irreversible curve indicates two types of problem. The obvious problem is that the overpotential of one redox couple may smear into the voltammogram of another, complicating the analysis by compromising resolution. A less obvious problem has to do with the possibility of associated homogeneous (solution) kinetics. Assume one wished to use the first voltammogram to determine the concentration of the first redox couple but the species giving the irreversible voltammogram is present in the sample. A likely choice of potential would be just before the cathodic limiting current, $i_{1,c}$, because the irreversible component is not reduced there at the surface of the electrode. Unfortunately, this results in a catalytic current or what electrochemists call an EC reaction. Just because the reduction of a species may be irreversible at an electrode surface does not mean that its reduction does not occur in solution. In the present case, the oxidized species that can be reduced electrochemically at the electrode surface may reduce the irreversible species in a following chemical reaction and thus be regenerated for further reduction at the electrode. This electrochemical reduction is the negative of the redox potential of the irreversible species.

Examples of such irreversible species (12) include hydroxylamine, hydroxide, and perchlorate. The electrochemistries of dichromate and thiosulfate are also irreversible. The presence of any of these agents may compromise an analysis by generating currents in excess of the analytically useful values. This problem can be avoided if the chemical reaction is slow enough, or if the electrode can be rotated fast enough so that the reaction does not occur within the Nernst diffusion layer and therefore does not influence the current.

Not all electroanalytical problems are confined to the measurement of currents. The presence of metal complexing agents has a profound effect on the selection of electrolysis potentials (see CHELATING AGENTS; COORDINATION COMPOUNDS). Ligands decrease the activity of at least one of the redox forms. Consider the polarographic reduction of a metal ion, M^{+n} , at a DME in the presence of a complexing ligand, L, assuming reversibility, so the Nernst equation is obeyed. M and L obey the equilibrium,



with an instability constant, K . Equation 5 is obeyed but the half-wave potential, given by equation 6, becomes (12)

$$E_{1/2}^C = E_{1/2}^M + \frac{RT}{nF} \ln K - p \frac{RT}{nF} \ln L + \frac{RT}{nF} \ln \left(\frac{m_M}{m_C} \right) \tag{7}$$

where $E_{1/2}^C$ and $E_{1/2}^M$ are the half-wave potentials of the complex and free metal, respectively, and m_M and m_C are the corresponding mass-transfer coefficients. The results of Equation 7, which can be used under favorable circumstances to determine p and K , may be a blessing or a curse. It is a curse if the intention is to measure the metal-ion concentration, because the presence of the ligand shifts the half-wave potential cathodically and the complex is harder to reduce. On the other hand, the addition of a specific ligand to the sample solution may remove interferents by shifting the half-wave potentials beyond that required for the analyte.

Voltammograms such as those shown in Figure 6 can be considered static because a steady-state current is measured at any given voltage. Cyclic voltammetry, on the other hand, is a dynamic technique because currents depend not only on applied voltages but also on the way the voltages are applied (sweep rates). Voltages are applied linearly as a function of time in one direction and then reversed or cycled back in the other direction. The result is a duck-shaped response, like curve A in Figure 7, for the reduction of a Ni(II) complex (14). As the potential is swept negatively from -0.2 volts, referenced against a saturated cadmium amalgam reference electrode, at 100 mV s^{-1} , the current initially rises because the electroactive species is reduced. The current decays, in accordance with equation 1, because the potential is so negative that the surface concentration of the species is held at zero and the current becomes diffusion controlled. The potential is reversed at -1.2 volts and swept in the opposite direction, resulting in the reoxidation of the reduced form of the species that remains near the surface of the electrode. The solution is not stirred or disturbed in any way.

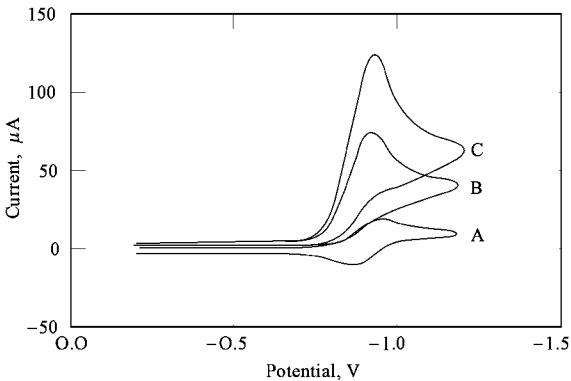
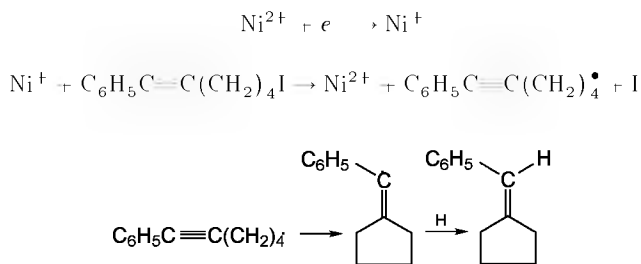


Fig. 7. Cyclic voltammograms for the reduction of 1.0 mM [2,2'-ethylene-bis(nitrilomethylidene)diphenolato]nickel(II) in dimethyl formamide at a glassy carbon electrode, in A, the absence, and B and C the presence of 2.0 and 5.0 mM 6-iodo-1-phenyl-1-hexyne, respectively (14).

In the presence of 6-iodo-1-phenyl-1-hexyne, the current increases in the cathodic (negative potential going) direction because the hexyne catalytically regenerates the nickel(II) complex. The absence of the nickel(I) complex precludes an anodic wave upon reversal of the sweep direction; there is nothing to reduce. If the catalytic process were slow enough it would be possible to recover the anodic wave by increasing the sweep rate to a value so fast that the reduced species (the nickel(I) complex) would be reoxidized before it could react with the hexyne. A quantitative treatment of the data, collected at several sweep rates, could then be used to calculate the rate constant for the catalytic reaction at the electrode surface. Such rate constants may be substantially different from those measured in the bulk of the solution. The chemical and electrochemical reactions involved are



It is important to understand the fundamental electrochemistries of analytes before attempting electroanalysis. The usual approach is to perform electroanalyses so quickly that kinetic events do not have time to occur before charge-transfer (electrolysis) has provided a response that is unequivocally related to the concentration of the analyte. Pulse techniques figure prominently into this principle. See Reference 10 for a highly useful approach to this problem.

Active electrochemical techniques are not confined to pulse and linear sweep waveforms, which are considered large amplitude methods. A-C voltammetry, considered a small amplitude method because an alternating voltage <10 mV is applied to actively couple through the double-layer capacitance, can also be used (15). An excellent source of additional information concerning active electroanalytical techniques can be found in References 16–18. Reference 18, although directed toward clinical chemistry and medicine, also contains an excellent review of electroanalytical techniques (see also AUTOMATED INSTRUMENTATION, CLINICAL CHEMISTRY).

Passive Techniques

Perhaps the most precise, reliable, accurate, convenient, selective, inexpensive, and commercially successful electroanalytical techniques are the passive techniques, which include only potentiometry and use of ion-selective electrodes, either directly or in potentiometric titrations. Whereas these techniques receive only cursory or no treatment in electrochemistry textbooks, the subject is regularly reviewed and treated (19–22). Reference 22 is especially recommended for novices in the field. Additionally, there is a journal, *Ion-Selective Electrode Reviews*, devoted solely to the use of ion-selective electrodes.

According to the definition, a passive technique is one for which no applied signal is required to measure a response that is analytically useful. Only the potential (the equilibrium potential) corresponding to zero current is measured. Because no current flows, the auxiliary electrode is no longer needed. The two-electrode system, where the working electrode may or not be an ion-selective electrode, suffices.

The equilibrium potential may be forced to change by the addition of a titrant, as in potentiometric titrations employing nonspecific working electrodes. But equilibrium potential can also be analytically useful in its own right if the response of the electrode is highly selective to the analyte. Chemically modified electrodes (23) are being researched that may provide this high selectivity, eg, for use in active techniques, such as the glucose sensor (24). However, ion-selective electrodes, of which the pH glass electrode is probably the most common example, dominate applications involving high selectivity even though their response mechanisms have nothing to do with oxidation and reduction. Potentiometry employing ion-selective electrodes without the addition of titrants is termed direct potentiometry.

Direct Potentiometry and Ion-Selective Electrodes. The reference electrode, Ag–AgCl, illustrated in Figure 2 may be taken as representative of reference electrodes in general. The difference in potential between this electrode and a working electrode forces a current flow that eventually discharges the cell like a battery if the two are shorted together. In practice, however, little or no current is allowed to flow during a potentiometric measurement because of the essentially infinite input impedance of the pH or millivolt meter used to measure the potential difference between them. It is convenient, therefore, to think of the current that would flow as a virtual current. The question then becomes one of what might carry virtual current. Electrons would obviously carry current through the external circuit. In the case of the Ag–AgCl electrode an electron that enters the reference electrode would reduce AgCl to free Ag⁰ and release Cl[–] into the adjacent (reference) solution. Conversely, an electron removed from the reference electrode would result in the oxidation of Ag⁰ to Ag⁺, which would precipitate Cl[–] from the reference solution as AgCl. In either case Cl[–] carries the current across the electrode–reference solution interface.

Current flow through the frits is supported by ions. Cations and anions both support the virtual current by flowing in opposite directions, and the transference number of a particular ion is defined as the fraction of the total current it carries. The sum of all transference numbers then is necessarily unity. If the fraction of the virtual current carried by the cations equals the fraction carried by the anions then the solution is said to be equitransferent.

It is important that a junction of equitransferent ions be present between the sample and reference solutions in order to minimize junction potentials, which result when ions of opposite sign are diffusing in the same direction. The free (very slowly) flowing junctions at the frits in Figure 2 can generate junction potentials if the anions and cations have unequal mobilities. The faster ion tends to get ahead of the slower one, causing the generation of a potential that slows down the fast one and speeds up the slow one until the two move with equal velocities, thus ensuring electroneutrality. Ions are equitransferent if the mobilities are equal. Thus junction potentials are minimized by equitransferent ions. Minimized generally means variation between samples of <0.25 mV, because a 0.25 mV error in potential measurement corresponds to a 1% error in predicted concentration (activity). This value can be derived by taking the derivative of the Nernst equation and allowing 100 da a % error, where a is the activity of the potential determining ion.

An ion-selective electrode, shown schematically in Figure 8, is constructed similarly to the reference electrode shown in Figure 2. Basically, the equitransferent frit has been replaced by an ion-selective membrane and the KNO_3 solution has been eliminated. This membrane is permselective if ions of only one sign can carry the virtual current. If only one particular ion can carry the virtual current then the electrode is said to be specific to that ion. Virtually all electrodes are selective, not specific, meaning that a number of ions having the same signs are sensed, but not with equal facility. Selectivity coefficients determine the degree to which ion-selective electrodes may generate potentials in response to the presence of different ions. Simple thermodynamic arguments reveal that these potentials are Nernstian such that there are 59 $\frac{\text{mV}}{z}$ millivolts per decade change in activity at room temperature if only one ion of charge z is being sensed. If several ions are sensed, the generally obeyed equation is the Nicholskii equation,

$$E - E^{0'} + \frac{RT}{F} \ln \left(a_i + \sum_j K_{i,j} a_j \right) \quad (8)$$

where a_i is the activity of the analyte ion, K_{ij} is the selectivity coefficient of the i^{th} ion in the presence of the j^{th} interferent, and a_j is the activity of the j^{th} interferent (20).

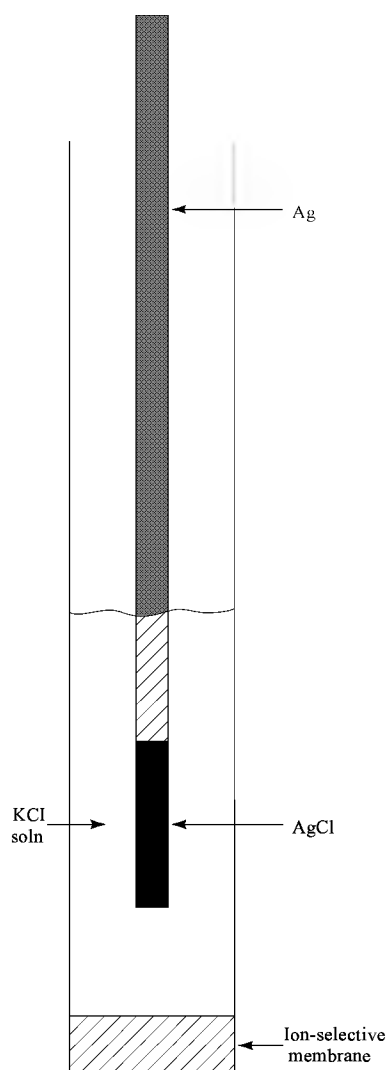


Fig. 8. Schematic of an ion-selective electrode.

Although the theory of ion-selective electrode responses has been presented in terms of transference, the selectivities of these electrodes have more to do with ion-exchange constants than with the mobilities of ions within their interiors (20,21). For example, protons do not carry virtual current through pH-selective glass membranes. Basically, an electrode is more selective to one ion than to another because the membrane can accommodate the presence of the one ion better than the other. Cavities in the membrane having just the right size, and perhaps shape, to accommodate a given ion can be envisioned. These cavities are completely filled with specific (analyte) ions where the charges are compensated for by sites of opposite sign that are confined within the membrane and, ideally, cannot enter the sample solution. These sites provide the membrane with permselectivity by electrostatically repelling ions of like sign. A potential difference is generated at one surface of the membrane when an ion is extracted from its interior into the sample solution, thus generating a charge separation across the double-layer capacitance and a corresponding potential. Samples more concentrated in the analyte tend to extract fewer analyte ions from the membrane, so the charge separation is less and the potential difference is lower. If the membrane is cation selective, its surface becomes less negative, and therefore more positive as the concentration of cations is increased in the sample solution. Large selectivity coefficients imply severely interfering ions that can extract analyte ions from the membrane by ion exchange, thus compromising the charge separation and rendering the interfacial potential difference less analytically useful.

Of interest is the manner in which cavities of the appropriate size are introduced into ion-selective membranes. These membranes typically consist of highly plasticized poly(vinyl chloride) (see MEMBRANE TECHNOLOGY). Plasticizers (qv) are organic solvents such as phthalates, sebacates, trimellitates, and organic phosphates of various kinds, and cavities may simply be the excluded volumes maintained by these solvent molecules themselves. More often, however, neutral carrier molecules (20) are added to the membrane. These molecules are shaped like donuts and have holes that have the same sizes as the ions of interest, eg, valinomycin [2001-95-8], $C_{54}H_{90}N O_{18}$, and nonactin [6833-84-7], $C_{40}H_{44}O_{12}$, or have wrap around structures like methyl monensin that have just the right length to legate selectively with the analyte. Although efforts are under way to bind these molecules to membranes, thus removing the neutral carrier feature, it has been pointed out repeatedly that the molecule must be able to physically carry the analyte ion between the two contacting solutions, ie, the sample and the reference (20,25,26).

Ion-selective electrodes are available for the electroanalysis of most small anions, eg, halides, sulfide, carbonate, nitrate, etc, and cations, eg, lithium, sodium, potassium, hydrogen, magnesium, calcium, etc, but having varying degrees of selectivity. The most successful uses of these electrodes involve process monitoring, eg, for pH, where precision beyond the unstable reference electrode's ability to deliver is not generally required, and for clinical applications, eg, sodium, potassium, chloride, and carbonate in blood, urine, and serum.

Ion-selective electrodes can also become sensors (qv) for gases such as carbon dioxide (qv), ammonia (qv), and hydrogen sulfide by isolating the gas in buffered solutions protected from the sample atmosphere by gas-permeable membranes. Typically, pH glass electrodes are used, but electrodes selective to carbonate or sulfide may be more selective.

Electrodes may also be rendered selective to more complex analytes using enzyme or other overcoats (see BIOPOLYMERS, ANALYTICAL TECHNIQUES; BIOSENSORS). The enzyme converts the analyte into a detectable ion or gas. Glucose and blood urea nitrogen sensors can be made in this way.

Potentiometric Titrations. If one wishes to analyze electroactive analytes that are not ions or for which ion-selective electrodes are not available, two problems arise. First, the working electrodes, such as silver, platinum, mercury, etc, are not selective. Second, metallic electrodes may exhibit mixed potentials, which may arise from a variety of causes. For example, silver may exchange electrons with redox couples in solution, sense Ag^+ via electron exchange with the external circuit, or tarnish to produce pH-sensitive oxide sites or Ag_2S sites that are sensitive to sulfide and halide. On the other hand, anodization of the silver electrode produces the very stable $Ag-AgCl$ electrode, which responds selectively to Cl^- , albeit interfered with by sulfide and

the other halides, for use not only in reference electrodes, but also for direct potentiometry.

In a titration the analytical utility of the measured potential lies not in its value, which may drift or be otherwise unstable, but in the magnitude of the change of its value near an end point. In a redox titration, the potential changes from something close to the E^0 of the analyte to something close to the E^0 of the titrant. This works fine provided the electrochemistries of both analyte and titrant are reversible. The technique may fail, however, if the electrode responds slowly to concentration changes because of irreversibility.

The Karl Fischer jack on the back of most pH meters, used to monitor Karl Fischer titrations, supplies a constant regulated current to the cell, which can consist of two identical (platinum) working electrodes. The voltammograms shown in Figure 9 illustrate the essential features of this technique. The sample corresponds to the wave on the right-hand (cathodic) side of each figure and is therefore easily oxidized. The titrant is represented by the wave on the left-hand (anodic) side and is therefore easily reduced. Halfway to the end point the potential difference, $\Delta E_{1/2}$, remains small, but at the end point the potential difference, ΔE_{ep} is large, nearly equal to the standard potential difference, ΔE^0 , between the sample and titrant redox couples. The shape of the titration curve depends on the irreversibilities of the analyte and titrant, but the large potential difference at the end point is analytically useful nevertheless. An example is the oxidation of Fe^{2+} by Ce^{4+} (27).

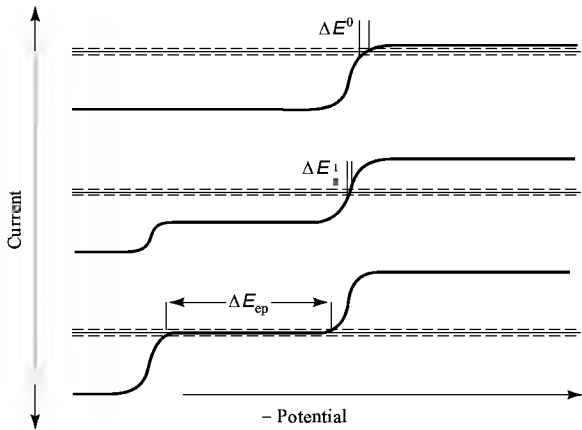


Fig. 9. Voltammograms demonstrating a potentiometric titration using dual-polarized electrodes, where the dashed lines indicate the anodic and equal-but-opposite cathodic currents that must be carried by the two opposing electrodes during the titration. Terms are defined in text.

Static and Dynamic Measurements

The definition herein of static electroanalytical measurements implies a time independent response, regardless of whether or not that response is generated by an applied signal. If an applied signal is needed, the method is active; if not, it is passive. These terms should not be confused with other definitions (1) where static more or less equates to passive and dynamic to active. This difference can be illustrated by several examples of dynamic measurements that can be either active or passive. First there is the batch injection analysis (BIA) technique (28). Basically, an electrode is placed upside down in a beaker of supporting electrolyte so that small aliquots of the sample can be injected onto its surface. A signal, measured during the residency of the sample, rapidly decays as the sample dissipates into the supporting electrolyte, which is stirred. Peak responses correlate well to sample concentrations. The term dynamic measurements is used for this technique. Both modified and unmodified carbon-paste electrodes were used amperometrically to demonstrate what are called active measurements herein. A silver–silver chloride ion-selective electrode is used to illustrate a passive approach.

BIA is closely related to flow injection analysis (FIA), where the sample is injected into a flowing stream of supporting electrolyte that carries it past the stationary working electrode. A transient response is again generated, regardless of whether or not the method is active or passive. Thus FIA is another example of a dynamic technique (29). Figure 10 illustrates results from this technique (29). Errors from ionic interferences can be eliminated in enzyme-based, flow-injection biosensing by the intentional addition of interferences to the sample. In this case the enzyme, glutaminase, was used to generate NH_4^+ from glutamine (the analyte) in a "split-stream" FIA system. Two peaks are generated for each injection because the two streams provide sample to the detector at different times. One stream provides only the excess NH_4^+ . The other passes through an enzyme reactor and therefore provides NH_4^+ generated by the glutamine as well. The presence of the interferent in relatively high concentration not only swamps out the effects of other interferences, but also linearizes the Nemst equation. The peaks in Figure 10a are only a few millivolts high. The peak corresponding to 0.5 mM glutamine (Fig. 10b) has an amplitude of only about 0.4 mV. Such small amplitudes may be cause for concern because the junction potential generated by the reference electrode can itself be several tenths of a millivolt.

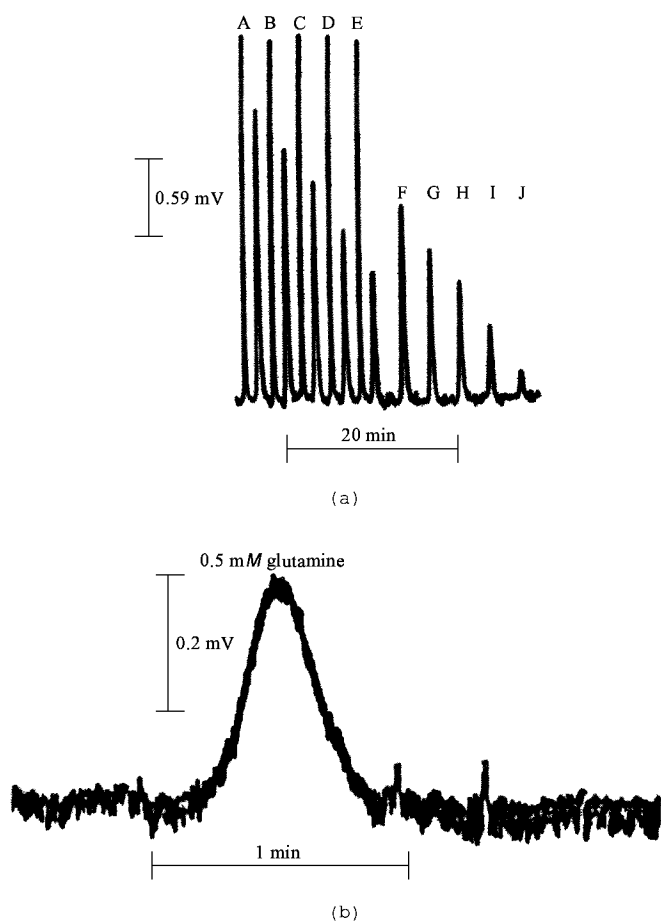


Fig. 10. Response of an ammonium ion-selective electrode which was made from nonactin dissolved in a PVC membrane mounted in a Phillips electrode body (ISE-561, Glasblaserie Müller, Zurich) referenced against an SCE to a split-stream FIA system. (a) FIA response to injected standards containing both glutamine and ammonium chloride: lines A–E contain both glutamine and ammonium chloride; F–G contain only glutamine. (b) Strip-chart recording illustrating signal to noise (S/N) for a 0.5 mM glutamine standard (29).

The duration of the response results primarily from the rate of elution of the sample, and not on any inherent limitation in the response time of the electrode. This is a characteristic of ion-selective electrodes, but amperometric responses depend not only on the duration of elution but also on flow rate because of the hydrodynamic effects discussed previously.

Another dynamic measurement is the LCEC technique which can be thought of, simplistically, as FIA using a chromatographic column positioned between the sample injection port and the detector. Bioanalytical systems (BAS) of West Lafayette, Indiana, specializes in instrumentation for LCEC. Their catalogs come with extensive bibliographies covering a variety of applications.

Economic Aspects

The reference electrode contributes heavily to the economics of electroanalytical chemistry. Companies that sell and service electroanalytical instrumentation are few in number and small in size, or they are parts of much larger companies. One supplier of electroanalytical instrumentation is Princeton Applied Research Corp. (PARC) of Princeton, New Jersey. PARC is a subsidiary of EG&G Instruments, Inc. Among the many suppliers of ion-selective electrodes are Orion (Boston, Massachusetts), Corning (Corning, New York), and Ingold (Wilmington, Massachusetts). Brinkmann Instruments, Inc. (Westbury, New York) is a useful supplier of titration equipment.

Electroanalytical chemistry does not yet generate the kind of revenue that battery technology and electrowinning (see METTALURGY, EXTRACTIVE) do. Electroanalysis, even at the most sensitive limits, can be performed using simple instruments that cost less than \$100. However, assays generally require highly skilled technicians and are thus correspondingly expensive even if the equipment used is not.

In the area of consumer products, amperometric glucose sensors hold high potential. Industrially, process monitors for the manufacture of consumer chemicals are under development. However, replacement of defective reference electrodes, which in a laboratory environment may be trivial, may be prohibitively difficult *in vivo* or in an industrial process environment.

BIBLIOGRAPHY

"Analytical Methods" in *ECT* 3rd ed., Vol. 2, pp. 587–683, by E. Lifshin and E. A. Williams, General Electric Co.; in *ECT* 3rd ed., Suppl., pp. 43–64, by C. W. Reimann and R. A. Velapoldi, National Bureau of Standards; "Ion-Selective Electrodes" in *ECT* 3rd ed., Vol. 13, pp. 720–734, by M. S. Frant, Foxboro Analytical.

1. P. T. Kissinger and W. R. Heineman, eds., *Laboratory Techniques in Electroanalytical Chemistry*, Marcel Dekker, New York, 1984.
2. R. W. Murray, in B. W. Rossiter and J. F. Hamilton, eds., *Physical Methods of Chemistry*, Vol. II, John Wiley & Sons, Inc., New York, 1986, Chapt. 6.
3. Z. Galus, in Ref. 2, Chapt. 3.
4. E. R. Brown and J. R. Sandifer, in Ref. 2, Chapt. 4.
5. J. E. Harrar, "Techniques, Apparatus, and Analytical Applications of Controlled-Potential Coulometry," in A. J. Bard, ed., *Electroanalytical Chemistry*, Vol. 8, Marcel Dekker, New York, 1975.
6. A. A. Vlček, J. Volke, L. Pospíšil, and R. Kalvoda, in Ref. 2, Chapt. 9.
7. E. Barendrecht, "Stripping Voltammetry," in Ref. 5, Vol. 2, 1967.
8. G. J. Patriarche, M. Chateau-Gosselin, J. L. Vandenbalck, and P. Zuman, "Polarography and Related Electroanalytical Techniques in Pharmacy and Pharmacology," in Ref. 5, Vol. 11, 1979.

9.

H. Siegerman, "Polarography of Antibiotics and Antibacterial Agents," in Ref. 8.

10.

J. Osteryoung and J. J. O'Dea, "Square-Wave Voltammetry," in Ref. 5, Vol. 14, 1986.

11.

P. Delahay, *Double Layer and Electrode Kinetics*, John Wiley & Sons, Inc., New York, 1965.

12.

A. J. Bard and L. R. Faulkner, *Electrochemical Methods, Fundamentals and Applications*, John Wiley and Sons, Inc., New York, 1980.

13.

R. M. Wightman and D. O. Wipf, "Voltammetry at Ultramicroelectrodes," in Ref. 5, Vol. 15, 1989.

14.

M. S. Mubarak and D. G. Peters, *J. Electroanal. Chem.* **332**, 127 (1992).

15.

D. E. Smith, "AC Polarography and Related Techniques: Theory and Practice," in Ref. 5, Vol. 1, 1966.

16.

C. N. Reilley and R. W. Murray, "Introduction to Electrochemical Techniques," in I. M. Kolthoff and P. J. Elving, eds., *Treatise on Analytical Chemistry*, Part I, Vol. 4, 1966, Chapt. 43.

17.

Ref. 12, Chapt. 1.

18.

J. Wang, *Electroanalytical Techniques in Clinical Chemistry and Laboratory Medicine*, VCH, New York, 1988.

19.

T. R. Kissel, in Ref. 2, Chapt. 2.

20.

W. E. Morf, *The Principles of Ion-Selective Electrodes and of Membrane Transport*, Elsevier, New York, 1981.

21.

H. Freiser, ed., *Ion-Selective Electrodes in Analytical Chemistry*, Vol. 2, Plenum Press, New York, 1980.

22.

A. Evans, in A. M. James, ed., *Potentiometry and Ion-Selective Electrodes*, John Wiley & Sons, Inc., New York, 1987.

23.

R. W. Murray, "Chemically Modified Electrodes," in Ref. 5, Vol. 13, 1984.

24.

W. J. Aston, "The Construction of Mediated Amperometric Biosensors," in A. P. F. Turner, I. Karube, and G. S. Wilson, eds., *Biosensors, Fundamentals and Applications*, Oxford University Press, New York, 1987.

25.

J. R. Sandifer, *Anal. Chem.* **60**, 1553 (1988).

26.

J. R. Sandifer, *Anal. Chem.* **61**, 2341 (1989).

27.

C. N. Reilley and R. W. Murray, "Fundamentals of Electrode Processes," in I. M. Kolthoff and P. J. Elving, eds., *Treatise on Analytical Chemistry*, Part I, Vol. 4, 1966, Chapt. 42.

28.

J. Wang and Z. Taha, *Anal. Chem.* **63**, 1053 (1991).

29.

W. Matuszewski, S. A. Rosario, and M. E. Meyerhoff, *Anal. Chem.* **63**, 1906 (1991).

James R. Sandifer
Eastman Kodak Company

ELECTROCHEMICAL MACHINING.

See ELECTROLYTIC MACHINING METHODS.

ELECTROCHEMICAL PROCESSING

Introduction,
Inorganic,
Organic,

Introduction

Electrochemical systems convert chemical and electrical energy through charge-transfer reactions. These reactions occur at the interface between two phases. Consequently, an electrochemical cell contains multiple phases, and surface phenomena are important. Electrochemical processes are sometimes divided into two categories: electrolytic, where energy is supplied to the system, eg, the electrolysis of water and the production of aluminum; and galvanic, where electrical energy is obtained from the system, eg, batteries (qv) and fuel cells (qv).

The industrial economy depends heavily on electrochemical processes. Electrochemical systems have inherent advantages such as ambient temperature operation, easily controlled reaction rates, and minimal environmental impact (qv). Electrosynthesis is used in a number of commercial processes. Batteries and fuel cells, used for the interconversion and storage of energy, are not limited by the Carnot efficiency of thermal devices. Corrosion, another electrochemical process, is estimated to cost hundreds of millions of dollars annually in the United States alone (see CORROSION AND CORROSION CONTROL). Electrochemical systems can be described using the fundamental principles of thermodynamics, kinetics, and transport phenomena.

Thermodynamics of Electrochemical Cells

Consider the cell



for which the electrode reactions are



and



The electrode where oxidation occurs is the anode; the electrode where reduction occurs is called the cathode. Electrons released at the anode travel through an external circuit and react at the cathode. The vertical lines denote phase separation; the squiggly lines separate a junction region. Although adjacent phases are in equilibrium, not all species are present in every phase. The membrane provides an ionic path for sodium ions that are transported from the anode to the cathode, but also separates the chlorine and hydrogen gases. Within the junction region, ie, the membrane, transport processes occur (see also ALKALI AND CHLORINE PRODUCTS; MEMBRANE TECHNOLOGY).

Determining the cell potential requires knowledge of the thermodynamic and transport properties of the system. The analysis of the thermodynamics of electrochemical systems is analogous to that of neutral systems. For ionic species, however, the electrochemical potential replaces the chemical potential (1).

$$\mu_i = RT \ln \lambda_i = RT \ln (c_i f_i a_i^\theta)$$

(3)

The electrochemical potential, μ_i , of a species is a function of the electrical state as well as temperature, pressure, and composition; λ_i is the absolute activity, which can be broken down into three parts as shown. For an electrolyte, A, which dissociates into ν_+ cations and ν_- anions, the chemical potential of the electrolyte can be expressed by

$$\mu_A = \nu_+ \mu_+ + \nu_- \mu_- = \nu RT \ln (c f_\pm a_\pm^\theta)$$

(4)

where the subscript $\pm$ indicates a mean coefficient. For instance,

$$f_\pm^\nu = f_+^{\nu_+} f_-^{\nu_-}$$

(5)

At open-circuit, the current in the cell is zero, and species in adjoining phases are in equilibrium. For example, the electrochemical potential of electrons in phases α and β are identical. Furthermore, the two electrochemical reactions are equilibrated. Thus,

$$2\mu_{\text{Cl}}^\epsilon = \mu_{\text{Cl}_2}^{\beta'} + 2\mu_e^{\alpha'}$$

(6)

and

$$2\mu_{\text{H}_2\text{O}}^\delta + 2\mu_e^\alpha = \mu_{\text{H}_2}^\beta + 2\mu_{\text{OH}}^\delta$$

(7)

The cell potential U is

$$FU = F(\Phi^\alpha - \Phi^{\alpha'}) = \mu_e^\alpha - \mu_e^{\alpha'}$$

(8)

By convention, U denotes the potential of the right electrode relative to the left. Substituting equations 6 and 7 gives

$$FU = \frac{1}{2}\mu_{\text{H}_2}^\beta - \mu_{\text{H}_2\text{O}}^\delta + \frac{1}{2}\mu_{\text{Cl}_2}^{\beta'} + (\mu_{\text{OH}}^\delta - \mu_{\text{Cl}}^\epsilon)$$

(9)

Because only differences in chemical potential can be measured, the chemical or electrochemical potential of each species is broken down as in equation 3. An arbitrary secondary reference state is defined for each compound. For instance, the chemical potential of chlorine gas is expressed as

$$\mu_{\text{Cl}_2} = \mu_{\text{Cl}_2}^\circ + RT \ln p_{\text{Cl}_2}$$

(10)

The superscript $^\circ$ refers to the ideal-gas secondary reference state, and p is the fugacity. Equation 4 is used for the ionic species. Hence,

$$FU = FU^\theta + RT \ln \left[\frac{p_{\text{H}_2}^{1/2} p_{\text{Cl}_2}^{1/2} c_{\text{OH}}^\delta}{a_{\text{H}_2\text{O}}^\delta c_{\text{Cl}}^\epsilon} \right] + 2RT \ln \left[\frac{f_{\text{NaOH}}^\delta}{f_{\text{NaCl}}^\epsilon} \right] + (\Phi^\epsilon - \Phi^\delta)$$

(11)

where the electric potential Φ of the solution has been defined by

$$\mu_{\text{Na}^+} = RT \ln c_{\text{Na}^+} + F\Phi$$

(12)

The standard cell potential U^θ is given by

$$FU^\theta = \frac{1}{2} \mu_{\text{H}_2}^\circ - \mu_{\text{H}_2\text{O}}^0 + \frac{1}{2} \mu_{\text{Cl}_2}^\circ + \left(\mu_{\text{OH}}^\theta - \mu_{\text{Cl}}^\theta \right)$$

(13)

In equation 11 the superscripts denoting the phases have been omitted where the meaning is clear, that is, for the fugacity of hydrogen and chlorine. U^θ is the composition independent part of the cell potential. The value of U^θ , 2.1875 V, can be obtained by looking up the values of the secondary reference-state quantities in a table of chemical thermodynamic properties (qv). It is convenient to consider the reaction as the sum of two half-cell reactions. The standard cell potentials are tabulated (2), typically at 25°C, as half-cell reactions referenced to the hydrogen electrode, the potential of which is arbitrarily set to zero. This yields 1.3595 V for the chlorine electrode and -0.828 V for the hydrogen electrode in basic media; the difference being 2.1875 V. The second term on the right side of equation 11 accounts for the principal variation of the cell potential with composition of the reactants.

The third term on the right side of equation 11 is the corresponding ratio of the mean molar activity coefficients. The fourth and last term results from transport processes in the junction region, and cannot be determined from thermodynamics. This term is often small and is not treated here. Neglecting this junction region and activity corrections leads to

$$U \approx U^\theta + \frac{RT}{2F} \ln \frac{p_{\text{H}_2} p_{\text{Cl}_2} c_{\text{OH}}^2}{a_{\text{H}_2\text{O}}^2 c_{\text{Cl}}^2}$$

(14)

Equation 14 is a form of the Nernst equation. The overall chemical reaction for the passage of 2 Faradays of charge in the external circuit from right to left is



Each reactant and product appears in the Nernst equation raised to its stoichiometric power. Thermodynamic data for cell potentials have been compiled and graphed (3) as a function of pH. Such graphs are known as Pourbaix diagrams, and are valuable for the study of corrosion, electrodeposition, and other phenomena in aqueous solutions. From the above thermodynamic analysis, the cell potential can be related to the Gibbs energy change

$$\Delta G = -nFU$$

(16)

Thus, because the standard cell potential for reaction 15 is positive, the reaction proceeds spontaneously as written. Consequently, to produce chlorine and hydrogen gas, a potential must be applied to the cell that is greater than the open-circuit value. This then becomes an example of an electrolytic process.

Predicting the cell potential requires knowledge of thermodynamic properties and transport processes in the cell. Conversely, the measurement of cell potentials can be used to determine both thermodynamic and transport properties (4).

Kinetics and Interfacial Phenomena

The rate of an electrochemical process can be limited by kinetics and mass transfer. Before considering electrode kinetics, however, an examination of the nature of the interface between the electrode and the electrolyte, where electron-transfer reactions occur, is in order.

Because some substances may preferentially adsorb onto the surface of the electrode, the composition near the interface differs from that in the bulk solution. If the cell current is zero, there is no potential drop from ohmic resistance in the electrolyte or the electrodes. Yet from the thermodynamic analysis it is seen that there is a measurable cell potential. The question from where this potential arises can be answered by considering the interface.

At the interface between phases, there is a region called the electrical double layer where potential variation occurs. Figure 1 shows this region. Although the electrode and electrolyte are overall electrically neutral, the metal electrode may have a net charge near the surface. The solution layer closest to the electrode contains specifically adsorbed ions and is called the inner Helmholtz plane (IHP). Ions that are hydrated, generally the cations, can approach the metal surface only to a finite distance, and comprise the outer Helmholtz plane (OHP). The nonspecifically adsorbed ions are distributed by thermal agitation; this region is called the diffuse double layer and lies just outside the OHP.

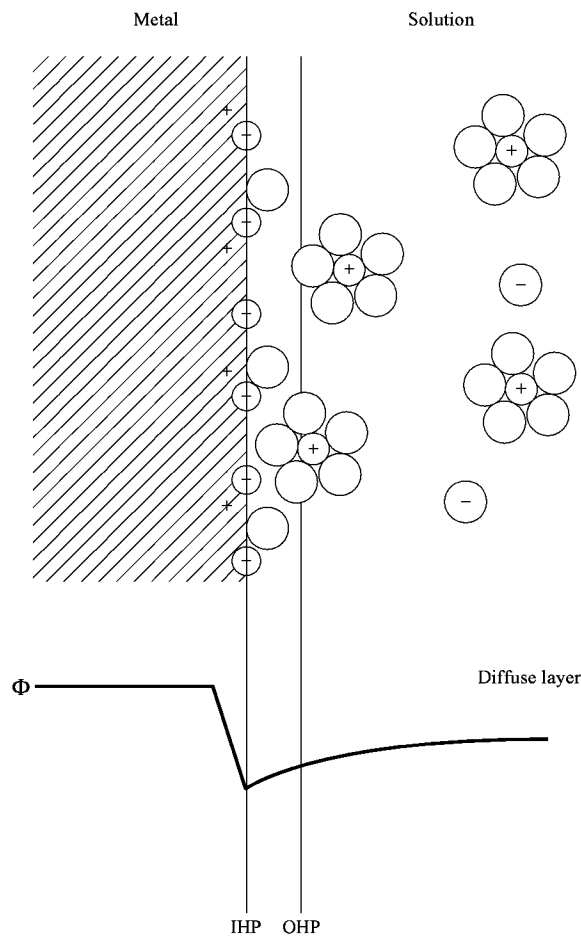


Fig. 1. The structure of the electrical double layer where $\bigcirc$ represents the solvent; $\ominus$, specifically adsorbed anions; $\ominus$, anions; and $\oplus$, cations. The inner Helmholtz plane (IHP) is the center of specifically adsorbed ions. The outer Helmholtz plane (OHP) is the closest point of approach for solvated cations or molecules. Φ , the corresponding electric potential across the double layer, is also shown.

In most electrochemical systems, the double layer is very thin (1–10 nm). The thickness is characterized by the debye length, λ ,

$$\lambda = \left(\frac{\epsilon RT}{2z^2 F^2 c_\infty} \right)^{1/2} \tag{17}$$

In addition, the potential of the electrode can be varied, resulting in a change in the structure of the interface. If no current is passed when the potential of the electrode changes, the electrode is called an ideally polarizable electrode, and can be described using thermodynamics.

Even in the absence of Faradaic current, ie, in the case of an ideally polarizable electrode, changing the potential of the electrode causes a transient current to flow, charging the double layer. The metal may have an excess charge near its surface to balance the charge of the specifically adsorbed ions. These two planes of charge separated by a small distance are analogous to a capacitor. Thus the electrode is analogous to a double-layer capacitance in parallel with a kinetic resistance.

In electrode kinetics a relationship is sought between the current density and the composition of the electrolyte, surface overpotential, and the electrode material. This microscopic description of the double layer indicates how structure and chemistry affect the rate of charge-transfer reactions. Generally in electrode kinetics the double layer is regarded as part of the interface, and a macroscopic relationship is sought. For the general reaction



the cathodic and anodic reaction rates can be written as

$$r_a = k_a c_R \exp \left[(1 - \beta) \frac{nF}{RT} V \right] \tag{19}$$

and

$$r_c = k_c c_O \exp \left[-\beta \frac{nF}{RT} V \right] \tag{20}$$

β is a symmetry factor equal to the fraction of the potential that promotes the cathodic reaction. The reaction rate and current are related through Faraday's law

$$\frac{i}{nF} = r_a - r_c \tag{21}$$

These two reactions may be generalized to the Butler-Volmer equation:

$$i = i_o \left\{ \exp \left(\frac{\alpha_a F \eta_s}{RT} \right) - \exp \left(-\frac{\alpha_c F \eta_s}{RT} \right) \right\} \tag{22}$$

The exchange current density, i_o , depends on temperature, the composition of the electrolyte adjacent to the electrode, and the electrode material. The exchange current density is a measure of the kinetic resistance. High values of i_o correspond to fast or reversible kinetics. The three parameters, α_a , α_c , i_o ,

are determined experimentally. The surface overpotential, η_s , is the difference in potential of the metal and the potential of an electrode of the same kind in the electrolyte measured adjacent to the electrode, ie, just outside the double layer, but passing no current. The surface overpotential appears in the exponential terms for both the anodic and cathodic reactions, and can be considered the driving force for the electrochemical reaction. The relationship between the surface overpotential and the current is shown in Figure 2.

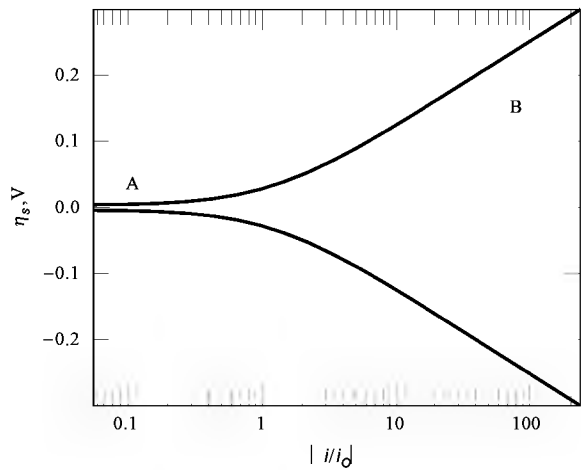


Fig. 2. Tafel plot, where α_a and α_c are both chosen to be 0.5; the temperature is 298.15 K. A indicates the linear region (eq. 23) and B the Tafel (eq. 24).

Two simplifications of the Butler-Volmer expression are often encountered. First, at low surface overpotentials equation 22 can be written as

$$i = (\alpha_a + \alpha_c) i_o \frac{F}{RT} \eta_s \tag{23}$$

This limit is called linear kinetics. On the other hand, if the surface overpotential is large, one of the exponential terms is negligible. This limit is called Tafel kinetics. The relationship was found empirically. In the anodic Tafel region

$$\eta_s = \frac{RT}{F \alpha_a} (\ln i - \ln i_o) \tag{24}$$

The Butler-Volmer equation can be applied to many, but not all, systems. Moreover, many of the systems that do not follow the Butler-Volmer model are of great practical importance, eg, in the corrosion of passivating metals (see CORROSION AND CORROSION CONTROL). Because of the unique nature of electron-transfer reactions, these have been of great theoretical interest. More recently, research has centered on a microscopic picture of the electron-transfer reactions and predicting reaction rate constants (5,6).

Transport Processes

In addition to electrode kinetics, the rate of an electrochemical reaction can be limited by the rate of mass transfer of reactants to and from the electrode surface. In dilute solutions, four principal equations are used. The flux of species i is

$$\mathbf{N}_i = -z_i u_i F c_i \nabla \Phi - D_i \nabla c_i + c_i \mathbf{v} \tag{25}$$

These three terms represent contributions to the flux from migration, diffusion, and convection, respectively. The bulk fluid velocity is determined from the equations of motion. Equation 25, with the convection term neglected, is frequently referred to as the Nernst-Planck equation. In systems containing charged species, ions experience a force from the electric field. This effect is called migration. The charge number of the ion is z_i , F is Faraday's constant, u_i is the ionic mobility, and Φ is the electric potential. The ionic mobility and the diffusion coefficient are related:

$$D_i = RT u_i \tag{26}$$

This relation, discovered by Nernst and Einstein, applies in the limit of infinite dilution.

A material balance on an element of the fluid gives

$$\frac{\partial c_i}{\partial t} = \nabla \cdot \mathbf{N}_i + R_i \tag{27}$$

where R_i is the homogeneous reaction rate. Except near the diffuse double layer, to a good approximation the solution is electrically neutral

$$\sum_i z_i c_i = 0 \tag{28}$$

The current density is given by

$$\mathbf{i} = F \sum_i z_i \mathbf{N}_i \tag{29}$$

These four equations, using the appropriate boundary conditions, can be solved to give current and potential distributions, and concentration profiles. Electrode kinetics would enter as part of the boundary conditions. The solution of these equations is not easy and often involves detailed numerical work.

Electroneutrality (eq. 28) is not strictly correct. More properly, equation 28 should be replaced with Poisson's equation

$$\nabla^2 \Phi = \frac{F}{\epsilon} \sum_i z_i c_i \tag{30}$$

Although the assumption of electroneutrality is appropriate for many systems, electroneutrality does not imply that Laplace's equation holds for the

potential. Further details are given in the literature (7).

When concentration gradients in the solution can be ignored, equations 25 through 29 show that the electric potential is governed by Laplace's equation

$$\nabla^2 \Phi = 0 \tag{31}$$

Laplace's equation is applicable to many electrochemical systems, and solutions are widely available (8). The current distribution is obtained from Ohm's law

$$\mathbf{i} = -\kappa \nabla \Phi \tag{32}$$

where κ is the conductivity of the solution. For dilute solutions the conductivity is given by

$$\kappa = F^2 \sum_i z_i^2 u_i c_i \tag{33}$$

Both the anions and cations can contribute to the current. In the absence of concentration gradients, the transference number relates the fraction of current carried by each species

$$t_j = \frac{z_j^2 u_j c_j}{\sum_i z_i^2 u_i c_i} \tag{34}$$

The distribution of current (local rate of reaction) on an electrode surface is important in many applications. When surface overpotentials can also be neglected, the resulting current distribution is called primary. Primary current distributions depend on geometry only and are often highly nonuniform. If electrode kinetics is also considered, Laplace's equation still applies but is subject to different boundary conditions. The resulting current distribution is called a secondary current distribution. Here, for linear kinetics the current distribution is characterized by the Wagner number, Wa , a dimensionless ratio of kinetic to ohmic resistance.

$$Wa = \frac{\kappa RT}{i_o (\alpha_a + \alpha_c) FL} \tag{35}$$

For large Wa , the current distribution is uniform. For example, when electroplating (qv) an object, usually a uniform deposit is desirable. Equation 35 suggests that a larger piece, ie, low Wa , would be more difficult to plate uniformly than a smaller one.

One other limit of interest is that as the rate of an electrochemical reaction increases (increasing surface overpotential), the concentration of reactants at the surface of the electrode decreases. When the concentration reaches zero, further changes in the overpotential do not increase the rate of reaction, ie, the reaction is mass-transfer limited. This condition is called limiting current. Current densities and concentration profiles for this condition are shown schematically in Figure 3. After plateauing, the current density often increases further in experimental systems as a result of the occurrence of a second Faradaic reaction at sufficiently high overpotential, eg, hydrogen evolution. This is shown in Figure 3a.

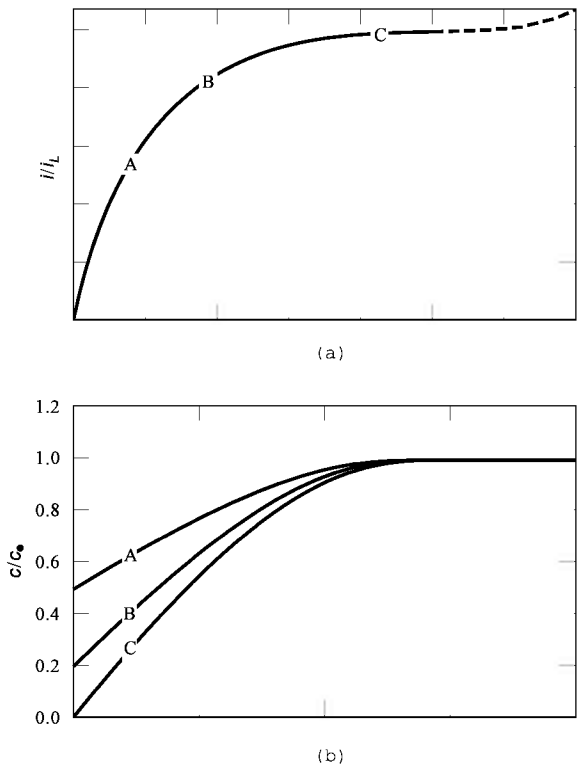


Fig. 3. Conditions of limiting current. (a) Current density on the electrode surface as a function of surface potential, showing points A, B, and C. The dashed line is for a second Faradaic reaction. (b) The concentration profiles corresponding to points A, B, and C.

Under conditions of limiting current, the system can be analyzed using the traditional convective-diffusion equations. For example, the correlation for flow between two flat plates is

$$Nu(x) = 1.2325 \left(\frac{ReSc d_e}{x} \right)^{1/3} - \frac{s_i i(x) d_e}{n F D c_\infty} \tag{36}$$

where d_e , the equivalent diameter of the flow channel, is equal to $2b$ for the flat-plate geometry. A number of mass-transfer correlations for systems relevant to electrochemistry are available (9).

As the Nemst equation suggests, concentration variations in the electrolyte lead to potential differences between electrodes of the same kind. These potential differences are concentration polarizations or concentration overpotentials. Concentration polarizations can also affect the current distribution. Predicting these is considerably more difficult. If concentration gradients exist, equations 25 and 27 through 29 must generally be solved simultaneously.

Electrochemical systems are found in a number of industrial processes. In addition to the subsequent discussions of electrosynthesis, electrochemical techniques are used to measure transport and kinetic properties of systems (see ELECTROANALYTICAL TECHNIQUES); to provide energy (see BATTERIES; FUEL CELLS); and to produce materials (see ELECTROPLATING). Electrochemistry can also play a destructive role (see CORROSION AND CORROSION CONTROL). The fundamentals necessary to analyze most electrochemical systems have been presented. More details of the fundamentals of electrochemistry are contained in the general references.

Nomenclature		
Symbol	Definition	Units
a	activity	
c_i	concentration of species i	mol m ³
d_e	equivalent diameter of annulus	m
D_i	diffusion coefficient	m ² s
f_i	molar activity coefficient of species i	
F	Faraday's constant	96,487 C eq
G	Gibbs energy	J mol
h	distance between wall of flow channel	cm
i	current density	A m ²
i_o	exchange-current density	A m ²
k	reaction rate constant	
L	characteristic length	m
N_i	molar flux of species i	mol (m ² s)
n	number of electrons transferred in reaction	
Nu	Nusselt number	
p	fugacity or pressure	kPa
r	heterogeneous reaction rate	mol (m ² s)
R_i	homogeneous reaction rate for species i	mol (m ³ s)
R	universal gas constant	8.3143 J (molK)
Re	Reynolds number	
s_i	stoichiometric coefficient	
Sc	Schmidt number	
t	time	s
T	temperature	K
U	open-circuit potential	V
u_i	mobility of species i	m ² mol ⁻¹ (Js)
v	velocity	m s
V	cell potential	V
Wa	Wagner number	
x	distance from entrance of channel	m
z_e	charge number of species i	
α	transfer coefficient	
β	symmetry coefficient	
ϵ	permittivity	F m
η_s	surface overpotential	V
κ	electrical conductivity	S m
λ	debye length	m
λ_i	absolute activity of species i	
μ_i	electrochemical potential of species i	J mol
v	total number of moles from electrolyte dissociation	
ν_+	number of moles of cations from electrolyte dissociation	
ν_-	number of moles of anions from electrolyte dissociation	
Φ	electrical potential	V
<i>Subscripts</i>		
∞	far from the electrode surface	
O	oxidized species	
R	reduced species	
a	anodic	
c	cathodic	
L	limiting current	
$+$	cation	
	anion	
<i>Superscripts</i>		
$*$	ideal gas	
θ	secondary reference state at infinite dilution	

BIBLIOGRAPHY

"Electrochemistry" in *ECT* 1st ed., Vol. 5, pp. 495–549, by G. C. Akerlof, Mellon Institute of Industrial Research; in *ECT* 2nd ed., Vol. 7, pp. 784–841, by G. C. Akerlof, Princeton University; "Introduction" under "Electrochemical Processing" in *ECT* 3rd ed., Vol. 8, pp. 662–671, by P. M. Robertson, Eidgenussische Technische Hochschule.

- E. A. Guggenheim, *J. Phys. Chem.* **33**, 842 (1929).
- A. J. Bard, R. Parsons, and J. Jordan, eds., *Standard Potentials in Aqueous Solution*, IUPAC, Marcel Dekker, Inc., New York, 1985.

3. M. Pourbaix, *Atlas of Electrochemical Equilibria in Aqueous Solutions*, NACE, Houston, Tex., 1974.
4. W. H. Smyrl and J. Newman, *J. Phys. Chem.* **72**, 4660–4671 (1968).
5. R. A. Marcus, *Ann. Rev. Phys. Chem.* **15**, 155 (1964).
6. M. J. Weaver in R. G. Compton, ed., *Comprehensive Chemical Kinetics*, Vol. 27, Elsevier, Amsterdam, 1987, pp. 1–60.
7. J. Newman, *Electrochemical Systems*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1991, pp. 256–259.
8. A. C. West and J. Newman in B. E. Conway, J. O'M. Bockris, and R. E. White, eds., *Modern Aspects of Electrochemistry*, Vol. 23, 1992, pp. 101–147.
9. J. R. Selman and C. W. Tobias, *Adv. Chem. Eng.* **10**, 211 (1978).

General References

J. Newman, *Electrochemical Systems*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1991.

A. J. Bard and L. R. Faulkner, *Electrochemical Methods*, John Wiley & Sons, Inc., New York, 1980.

G. Prentice, *Electrochemical Engineering Principles*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1991.

K. J. Vetter, *Electrochemical Kinetics*, Academic Press, Inc., New York, 1967.

A. T. Kuhn, *Industrial Electrochemical Processes*, Elsevier Publishing Co., Amsterdam, 1971.

Series on Electrochemistry

C. W. Tobias and H. Gerischer, eds., *Advances in Electrochemical Science and Engineering*, VCH, New York, from 1990.

C. W. Tobias and P. Delahay, eds., *Advances in Electrochemistry and Electrochemical Engineering*, John Wiley & Sons, Inc., New York, 1961–1984.

B. E. Conway, J. O'M. Bockris, and R. E. White, eds., *Modern Aspects of Electrochemistry*, Plenum Press, New York, from 1954.

J. O'M. Bockris and co-workers, eds., *Comprehensive Treatise on Electrochemistry*, Plenum Press, New York, from 1980.

A. J. Bard, ed., *Electroanalytical Chemistry*, Marcel Dekker, Inc., New York, from 1966.

Thomas F. Fuller
John Newman
University of California, Berkeley

INORGANIC

Electrochemical processes have been used to produce reactive inorganic chemicals and metals since the early 1800s. Many reactive elements were first isolated by electrolysis. Electrochemical processes involve cell components, reactor designs, and operating principles that can be very process-specific. The amount of electric energy required depends on the electrode material and the species reacting at the electrode. Electrode potentials range up to 3.5 V. This moderate level of energy is sufficient to produce the strongest oxidizing and reducing agents known, fluorine and solvated electrons, respectively. Efficient chemical and electrochemical processes exist for the production of some chemicals. Process choice then depends on a number of factors including capital costs, operating costs, and health and environmental considerations. Environmental factors generally favor electrochemical processing. It would be difficult, if not impossible, to develop chemical manufacturing processes for such highly reactive chemicals as sodium [7440-23-5], Na (see SODIUM AND SODIUM ALLOYS) and fluorine (qv) [7782-41-4], F₂.

The electrochemical production of inorganic chemicals and metals in the United States consumes about 5% of all the electricity generated annually, and about 16% of the electric power consumed by industry. This includes the production of such commodity chemicals as sodium hydroxide [1310-73-2], NaOH, and chlorine [7782-50-2], Cl₂, ranked eighth and ninth in volume of production in 1991, respectively (1). Table 1 lists the most important inorganic products of electrochemical processes. Electroplating (qv), electropolishing (see METAL SURFACE TREATMENTS), anodizing, electrolytic machining, energy sources such as batteries (qv) and fuel cells (qv), and other industrial operations involving the use of electric power and electrode reactions are all discussed elsewhere.

Table 1. 1991 U.S. Production of Electrochemical Products^a

Product	Quantity, <i>t</i> × 10 ³	Reference	Price, ^b \$ t	Product value, \$ × 10 ⁶
aluminum	4100	3	1320	5.41
beryllium	0.17	3 ^c		
cadmium	1.6	3	2200	0.0035
caustic potash	273	4	~700	0.191
caustic soda	11,091	5	330	3.66
chlorine	10,727	6	137.50	1.475
chromium	420	3 ^c	~2860	1.2
copper	1,630	3	2618	4.267
lithium	2.8	3 ^c	69,300	0.194
magnesium	130	3	3146	0.063
manganese metal	20	7	1188	0.024
manganese dioxide	39	7	1540	0.06
nickel	8.4	3	7480	0.063
perchlorates	20	7	~1100	0.02
permanganates	21	8	2750	0.058
sodium	71.8	9 ^d	2530	0.182
sodium chlorate	315	4	456	0.144
titanium	16	3		
zinc	520	3	1430	0.744
<i>Total</i>				<i>17.76</i>

^a Fluorine and persulfates were also produced electrochemically.
^b All product values are based on prices quoted in Ref. 2.
^c Value represents consumption rather than production.
^d Value is production for 1990.

Hardware for Electrochemical Processing

Power supplies and electrolytic cells are the distinguishing features of electrochemical processes. Nearly all electric power is generated and transported as high voltage multiple-phase alternating current. Industrial electrochemical processes require direct current; thus transformers are needed to decrease voltage, and rectifiers are required to convert the alternating current to direct current. Rectifiers may be rated for voltages up to 400 V or more and for any amperage up to hundreds of thousands of amperes. Rectifier efficiencies generally increase as voltages increase. Very high amperages are achieved by connecting rectifier units in parallel. State-of-the-art rectifiers use thyristers which are semiconductor devices that conduct only when a triggering potential is applied (see SEMICONDUCTORS). Thyristers can be made to conduct for half of an alternating current cycle and not conduct the other half-cycle, rectifying alternating current to direct current.

Electrochemical processes require feedstock preparation for the electrolytic cells. Additionally, the electrolysis product usually requires further processing. This often involves additional equipment, as is demonstrated by the flow diagram shown in Figure 1 for a membrane chlor-alkali cell process (see ALKALI AND CHLORINE PRODUCTS). Only the electrolytic cells and components are discussed herein.

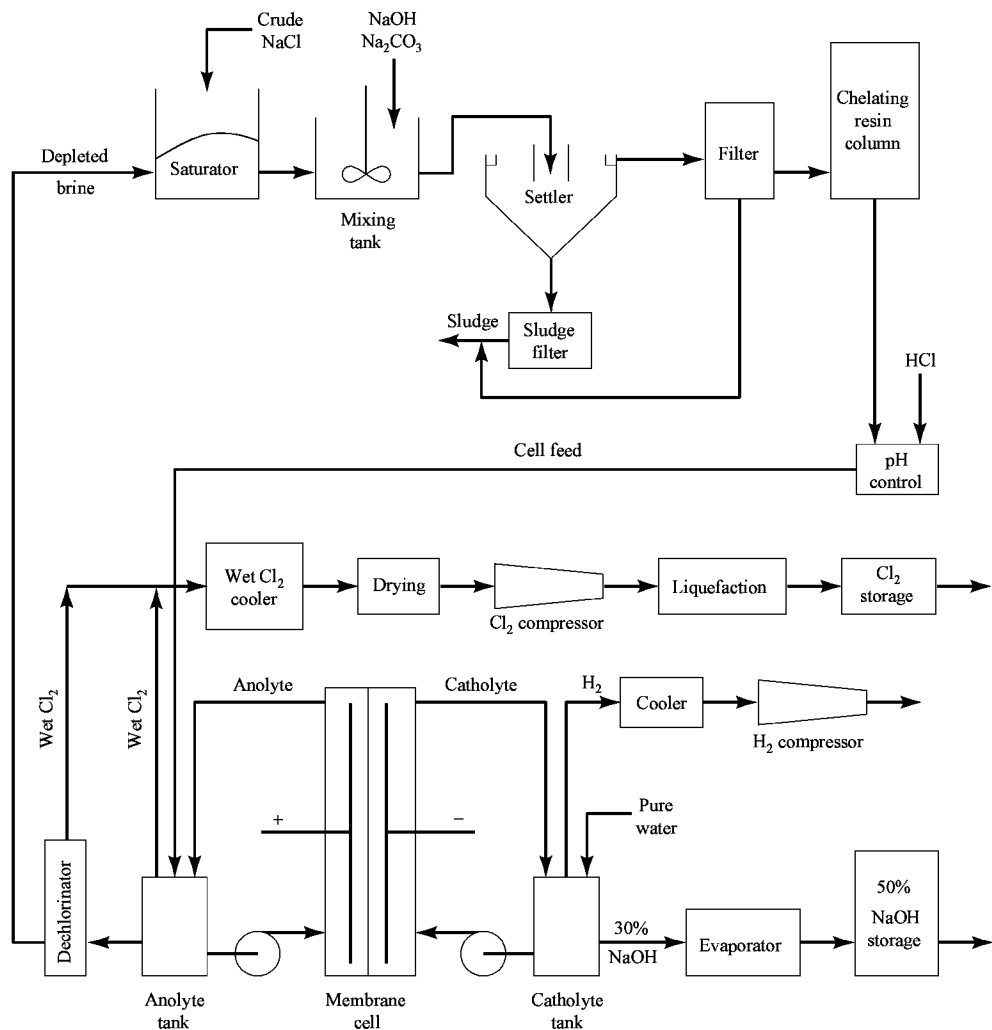


Fig. 1. Flow diagram for chlor-alkali production by a membrane cell process.

Design possibilities for electrolytic cells are numerous, and the design chosen for a particular electrochemical process depends on factors such as the need to separate anode and cathode reactants or products, the concentrations of feedstocks, desired subsequent chemical reactions of electrolysis products, transport of electroactive species to electrode surfaces, and electrode materials and shapes. Cells may be arranged in series and or parallel circuits. Some cell design possibilities for electrolytic cells are

Cell type	Process
one-compartment cells	
open-top inert tank	
monopolar electrodes	MnO ₂
bipolar electrodes	chlorate
open-top cathodic tank	chlorate and perchlorate
enclosed horizontal liquid metal cathode	chlorine-caustic and aluminum
two-compartment cells	
diaphragm cells	
monopolar electrodes	chlorine-caustic and Mn metal
bipolar electrodes	chlorine-caustic
membrane cells	
monopolar electrodes	chlorine-caustic
bipolar electrodes	chlorine-caustic
pressurized cells	
filter press	
bipolar electrodes	
rotating cylindrical cathode	hydrogen-oxygen
particle-bed electrodes	continuous production of metal
	removal of low concentration of metals from waste streams

Industrial electrolytic cells that produce gas at one or both electrodes are usually designed to take advantage of gas lift electrolyte circulation. Good electrolyte circulation provides transport of electroactive species to the electrode surfaces, and in some cases moves electrode products away from the electrode. Internal gas lift circulation in a well-designed cell is much more efficient than mechanically pumped circulation. The Chemetics sodium chlorate [775-09-9], NaClO₃, cell shown in Figure 2 is a good example of gas lift electrolyte circulation (see also CHLORINE OXYGEN ACIDS, CHLORIC ACID AND CHLORATES). Other examples are the Huron sodium chlorate cell shown in Figure 3, the Oxytech chlor-alkali membrane cell shown in Figure 4, the Lurgi membrane cell (Fig. 5), improved Hooker fluorine cell (Fig. 6), and a typical metal electrowinning cell (Fig. 7).

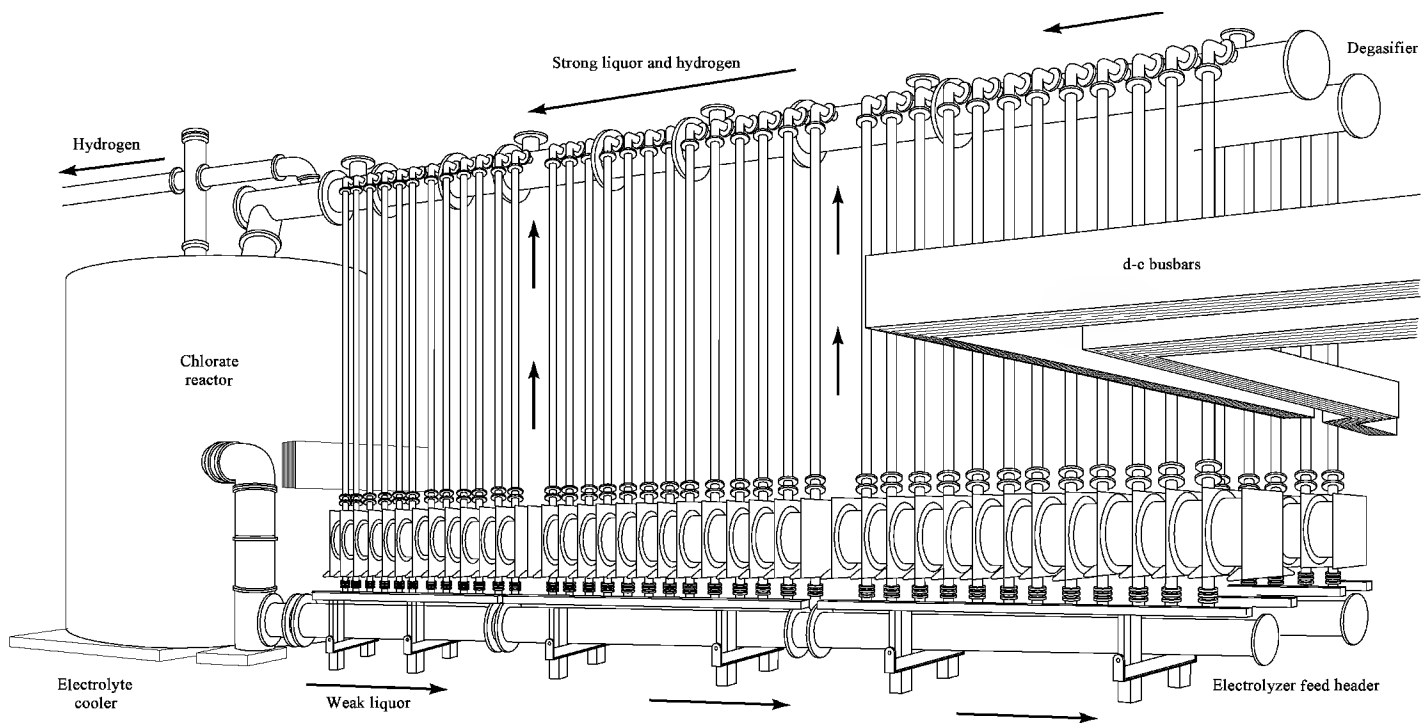


Fig. 2. Chemetics sodium chlorate cellroom.
Courtesy of Chemetics International Co., Ltd.

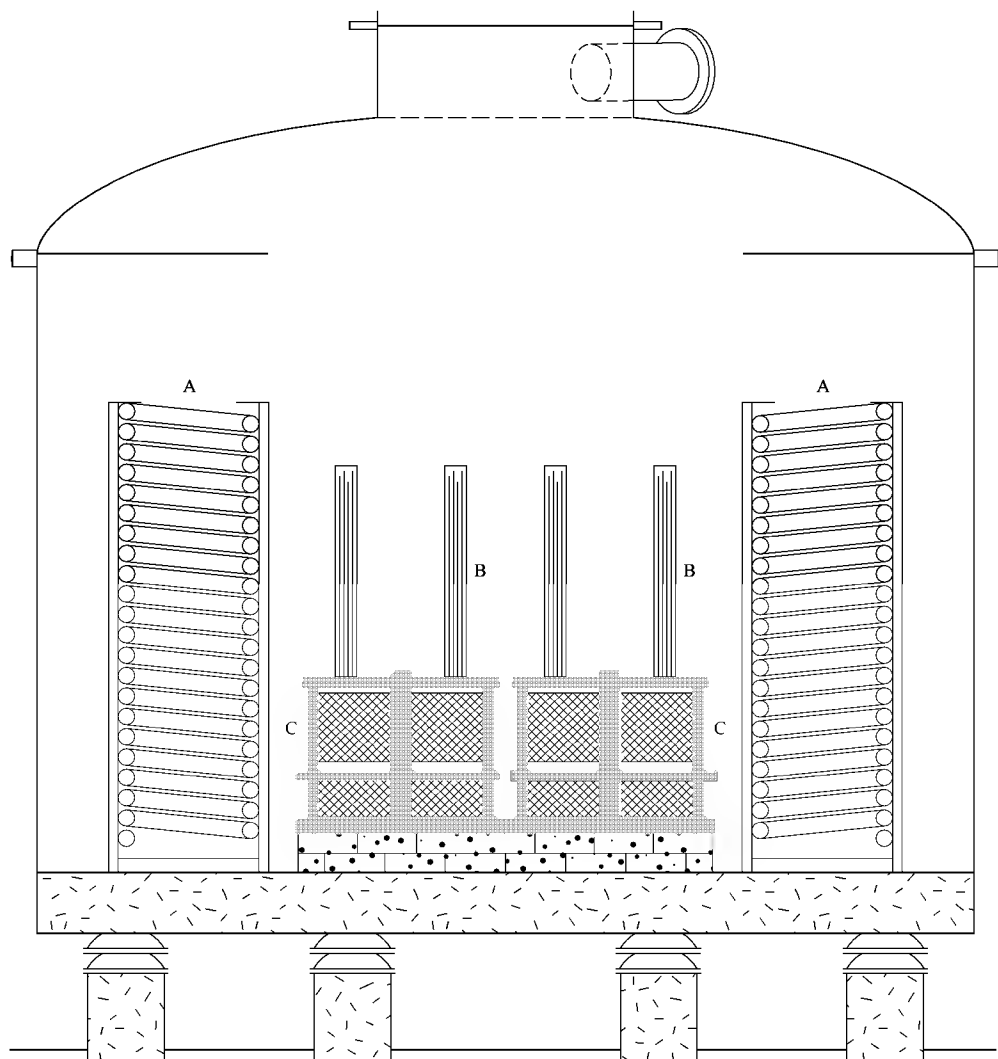


Fig. 3. Huron sodium chlorate cell system, where A corresponds to cooling coils; B to chimneys to promote gas lift circulation; and C to cell boxes. Electric bus connections are not shown.
Courtesy of Huron Technologies Inc.

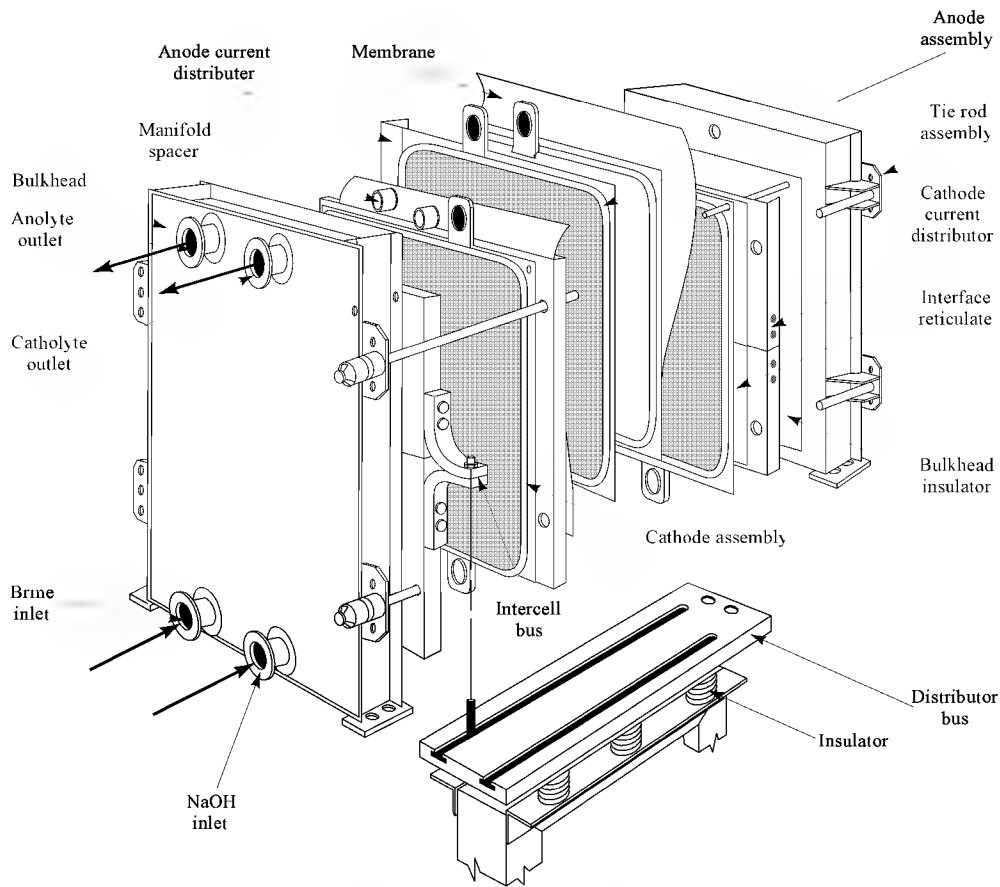


Fig. 4. OxyTech membrane gap cell electrolyzer assembly.

Courtesy of OxyTech Systems, Inc.

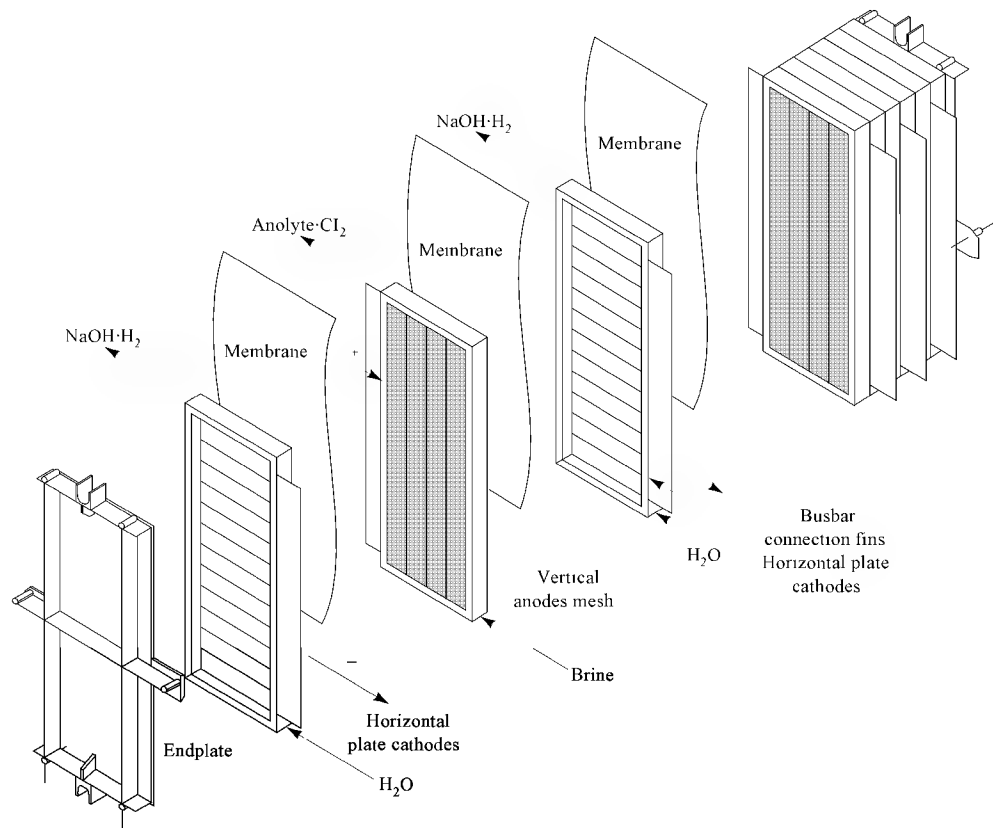


Fig. 5. Lurgi membrane electrolyzer.

Courtesy of Lurgi AG.

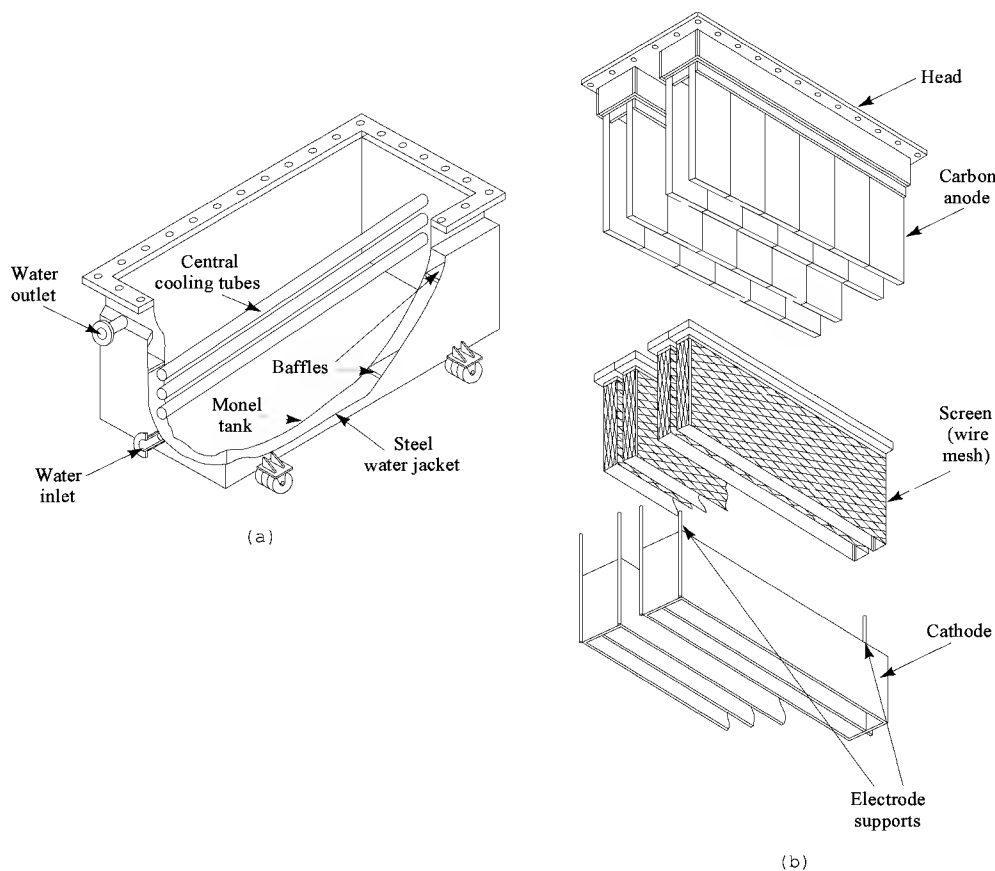


Fig. 6. Improved fluorine cell assembly: (a), cell tank; (b), electrode and diaphragm components.

Courtesy of American Chemical Society (10).

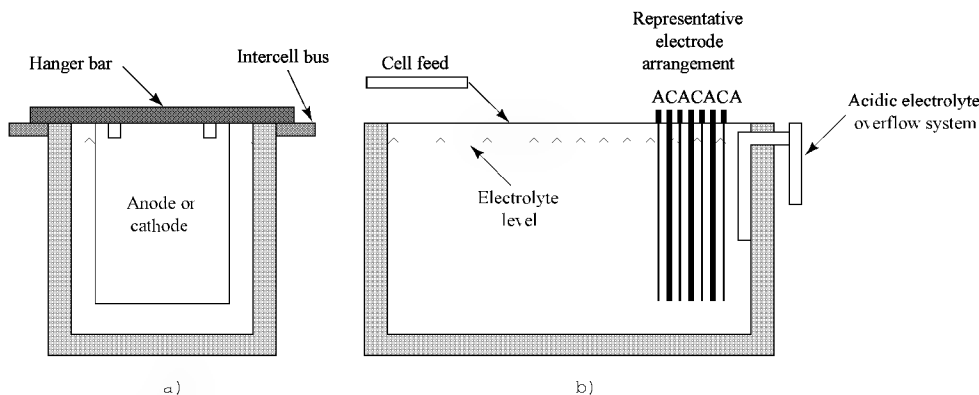


Fig. 7. Sections of metal electrowinning cell without diaphragm: (a), end view; (b), side view where A is anode and C is cathode.

The electrolytic cells shown in Figures 2–7 represent both monopolar and bipolar types. The Chemetics chlorate cell (Fig. 2) contains bipolar anode–cathode assemblies. The cathodes are Stahmet, a registered trademark of Chemetics International Co., and the anodes are titanium [7440-32-6], Ti, coated either with ruthenium dioxide [12036-10-1], RuO₂, or platinum [7440-06-4], Pt–iridium [7439-88-5], Ir (see METAL ANODES). Anodes and cathodes are joined to carrier plates of explosion-bonded titanium and Stahmet, respectively. Several individual cells electrically connected in series are associated with one reaction vessel.

Monopolar electrodes have a direct electrical connection with an external power supply. This requires the distribution of current over the total area of one monopolar electrode, collecting the current from the other monopolar electrode for conduction to the next cell through intercell busbars. Monopolar cells operate at low voltages, and may require high amperages. Industrial circuits of cells may consist of one hundred or more monopolar cells in series. Monopolar electrodes are used in some membrane chlor-alkali cells (Figs. 4 and 5), fluorine cells (Fig. 6), and in metal electrowinning cells (Fig. 7).

Bipolar electrodes have no direct electrical busbar connection to a power supply. Only the terminal electrodes that lead current into and out of a bank of cells are connected to a power source. These terminal electrodes are monopolar. Bipolar cells are polarized by potential gradients that force current to flow through the bipolar units. If conducting pathways around bipolar units are available, some current bypasses or leaks past bipolar units. Care must be taken when designing bipolar cells to minimize current leakage through cell feed manifolds and other possible current leakage pathways. Bipolar electrodes are polarized anodically on the side facing the negative terminal electrode, and cathodically on its reverse side. Bipolar cells may be flat side to side units or elongated units that are bipolar end to end (Figs. 2 and 3). Banks of cells are called electrolyzers. Electrolyzers may contain up to one hundred or more unit cells. The voltage required by a bipolar electrolyzer is the individual cell voltage multiplied by the number of cells in the electrolyzer. Bipolar cells generally require relatively low amperages. Large numbers of cells in electrolyzers and several electrolyzers in a circuit may require a relatively high rectifier voltage and relatively low amperage. Discussions of monopolar and bipolar cells, advantages and disadvantages, are found in References 11 and 12. Bipolar cell designs minimize energy consumption by operating at lower unit cell voltage.

The separation of anode and cathode products is often necessary. Electrolysis products are often very reactive and mixing can be hazardous. Examples are the electrolysis of water to produce oxygen [7782-44-7], O₂, at the anode and hydrogen [1333-74-0], H₂, at the cathode, and the electrolysis of sodium chloride [7647-14-5], NaCl, brine to produce chlorine [7782-50-5], Cl₂, and hydrogen. Cells for these electrochemical processes include a membrane or diaphragm to separate the anode and cathode compartments. Porous diaphragms made of deposited fibers, woven fabrics, or metallic wire mesh may be used in many cases. Porous diaphragms or ion-exchange (qv) membranes are used in water electrolysis cells and diaphragm-chlor-alkali cells (see MEMBRANE TECHNOLOGY). Woven nickel [7440-02-0], Ni, wire mesh diaphragms are used in fluorine cells (Fig. 6), and woven plastic fabrics are used in some metal

electrowinning cells. The latter are sometimes referred to as bags, and may be around the cathodes or anodes. A neutral cell feed (pH ~ 7) may enter the cathode compartment, percolate through the diaphragm, and become acidic anolyte.

Ion-exchange membranes are used when selective ion transport between compartments is desired and liquid flow through pores is undesirable. The chlor-alkali cells (Figs. 4 and 5) are examples of membrane cells. Very good ion-exchange membranes for electrolytic cells have been developed and are used commercially. Most newer chlorine and caustic plants are based on membrane cells, and several chlor-alkali plants have been converted to membrane cells (4). In chlor-alkali cells, cation-exchange membranes allow the transport of sodium ions from the anolyte compartment to the catholyte compartment, and nearly eliminate other ionic transport through the membrane. Membrane cells produce caustic that contains very little (ppm) sodium chloride. Commercial, ie, 50% o, caustic soda from diaphragm cells contains about 1% o sodium chloride.

Electrode materials and shapes may have a profound effect on cell designs. Anode materials encountered in electrochemical processes are

Anode material	Process
carbon	fluorine
	aluminum
graphite	chlorine–caustic
	chlorate
DSA	chlorine–caustic
	chlorate
Pt	perchlorates
	persulfates
	chlorate
lead dioxide, PbO ₂ , on graphite or titanium	perchlorate
lead alloy (Pb, 1% o Ag)	zinc electrowinning
	manganese electrowinning

Ideal anode materials should allow high anodic polarizations without being oxidized or deteriorated in any way. Depending on the process, some anode materials such as carbon [7440-44-0], C, graphite [7782-42-5], and lead dioxide [1309-60-0], PbO₂, corrode significantly and have relatively short useful lives. Hydrogen diffusion anodes similar to those used in fuel cells (qv) are being developed for use in Zn and Mn electrowinning cells (13).

Dimensionally stable anodes (DSA) are titanium anodes coated with ruthenium dioxide, RuO₂, and titanium dioxide, TiO₂ (14). DSA is a registered trademark of Eltech Systems Corp. DSA have replaced graphite anodes in nearly all commercial chlor-alkali and chlorate cells. The service life of DSA in diaphragm chlor-alkali cells has exceeded 12 years in many cases. DSA are produced by baking a catalyst layer, RuO₂–TiO₂, on a valve metal support (Ti,Ta). Other noble metal or noble metal oxide coatings on titanium are used in some applications. These may be platinum, platinum–iridium, platinum oxide, or ruthenium oxide in proprietary mixtures of catalyst, binder, and stabilizer. Activated electrodes offer superior corrosion resistance, lower reaction voltages, and lower electric power consumption per unit of production.

Cathode materials and electrochemical applications are

Cathode material	Process
mild steel	chlorine–caustic, chlorate, perchlorate
mercury	chlorine–caustic
nickel	fluorine
stainless steel	chlorine–caustic
active cathode coating	chlorine–caustic
O ₂ or air reduction	chlorine–caustic

Attempts to decrease the hydrogen overvoltage by applying active coatings of Ni–Zn, Ni–Al, and proprietary combinations of metals on steel, stainless steel, and nickel have met with some success (15). Cell voltages have been decreased by up to 0.2 V for more than 800 days (16,17). Oxygen or air cathodes similar to those used in fuel cells have been investigated for potential application in chlor-alkali cells, chlorate cells, and other cells producing hydrogen at the cathode (18). As of this writing, there are none used in inorganic electrochemical processing.

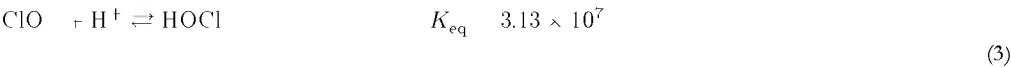
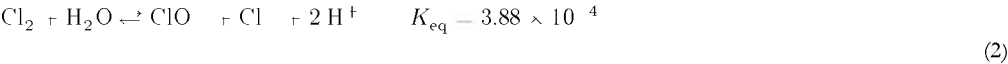
Mild steel cathodes are used extensively in chlor-alkali and chlorate cells. Newer activated cathode materials have been developed that decrease cell voltages about 0.2 V below that for cells having mild steel cathodes. Some activated cathodes have operated in production membrane cells for three years with only minor increases in voltage (17). Activated cathodes can decrease the energy consumption for chlorine–caustic production by 5 to 6.5% o.

Industrial Process Conditions

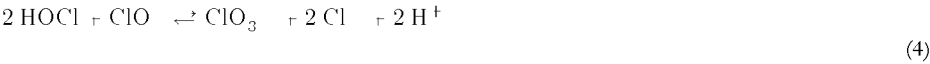
Electrolysis of Chloride Solutions. Chloride may be oxidized electrochemically to chlorine or hypochlorite, chlorate, and perchlorate. The distribution of products of the first oxidation step



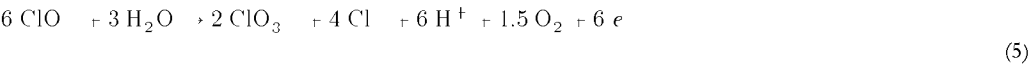
is very pH-dependent and is determined by the equilibria (19,20):



The distribution of species as a function of anolyte pH is shown in Figure 8 (21). Chlorine is the primary product at pH < 4. Hypochlorous acid [7790-92-3], HOCl, is the dominant species from pH 4 to pH 6. Above pH 6 combinations of HOCl and hypochlorite ion, ClO⁻, exist but react to produce ClO₃⁻ (22):



Hypochlorite ions also react at the anode to produce chlorate electrochemically. This reaction



is undesirable in commercial cells because it requires more electric power. The theoretical electric current required to produce chlorate by equations 1–4 is 6 Faraday per mole. Equations 1,2,3, and 5 require 9 Faradays per mole of chlorate produced. In commercial cells, reaction volume is provided to promote equation 4 and minimize equation 5.

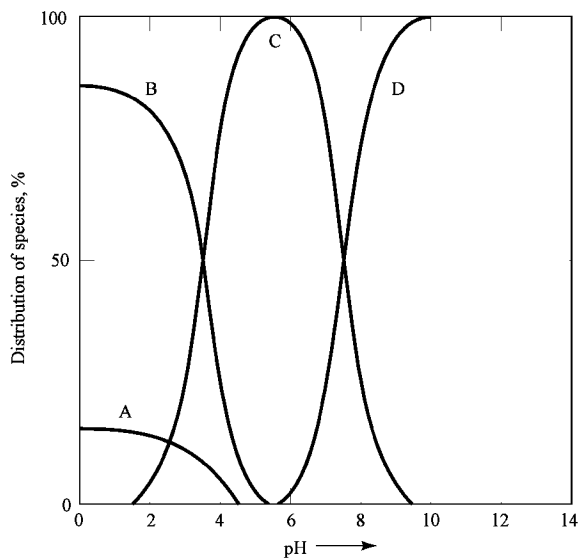


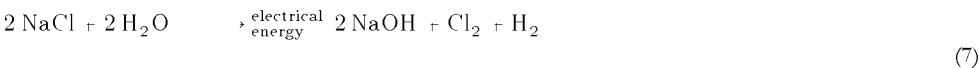
Fig. 8. Distribution of chlorine-containing species in a 1 M chloride solution. A, ClO₃⁻; B, Cl₂; C, HOCl; and D, ClO⁻.

The oxidation of chlorate to perchlorate, ClO₄⁻, is an electrochemical anode reaction independent of pH.



Perchlorate is usually produced by first producing chlorate, separating chlorate from chloride by crystallizing NaClO₃, and redissolving NaClO₃ to produce a nearly chloride-free chlorate solution for perchlorate production. Optimum electrolysis conditions and cell designs for chlorate and perchlorate production are different. Chlorate cells may use graphite, DSA, or PbO₂ anodes, are pH sensitive, and require a reaction volume to promote equation 4. Perchlorate cells may use platinum or PbO₂ anodes, are not pH sensitive, and require no reaction volume (see PERCHLORIC ACID AND PERCHLORATES).

The electrolysis of NaCl brine for the production of chlorine and caustic soda is one of the oldest and certainly one of the most important industrial electrochemical processes (22–26). The overall reaction is



There are three main technologies available for carrying out this process: diaphragm cells, mercury cells, and membrane cells. Membrane cells are the most recent development, and are generally chosen for new production capacity.

Diaphragm Cell Technology. Diaphragm cells feature a porous diaphragm that separates anode and cathode compartments of the cell. Diaphragms should provide resistance to liquid flow, require minimum space between anode and cathode, produce minimum electrical resistance, and be durable. At the anode, which is generally a DSA, chloride ions are oxidized to chlorine (see eq. 1) and at the cathode, which is usually a woven steel wire mesh, water is reduced to hydrogen.



The diaphragm, generally vacuum deposited on the cathode, separates the hydrogen and chlorine gases produced and the anolyte from the caustic soda containing catholyte. Purified brine containing about 26° NaCl is fed into the anode compartment. The electrolyte flows through the diaphragm into the cathode compartment and out of the cell as a 10–12° NaOH solution containing unreacted NaCl. This NaOH solution is evaporated to produce 50° NaOH. The solubility of NaCl decreases as NaOH concentrations increase in the evaporators, causing NaCl to precipitate. Solid NaCl is removed from the concentrated NaOH solution. Diaphragm cell caustic soda generally contains about 1° NaCl.

Diaphragm cells have been improved by using DSA instead of graphite anodes and by using modified asbestos (qv) diaphragms (27,28) and nonasbestos diaphragms (29). The operating characteristics of some diaphragm cells are given in Table 2.

Table 2. Operating Characteristics of Diaphragm Cells^a

Cell parameter	Oxytech				Glanor		Uhde	
model	H2A	H4	MDC29	MDC55	V-1144	V-1161	HU30 ^b	HU60 ^b
current rating, kA	80	150	80	150	72	72	60	120
current density, kA m ²	2.21	2.33	2.76	2.74	2.05	1.47		
cell voltage, V	3.35	3.40	3.60	3.59	3.50	3.08		
energy (d-c) consumed, kWh t ⁻¹	2655	2694	2847	2839	2800	2400		
cells electrolyzer	1	1	1	1	11	11	1	1
chlorine capacity, t d ⁻¹	2.42	4.54	2.43	4.55	24.2 ^c	24.2 ^c	1.82	3.64
caustic capacity, t d ⁻¹	2.73	5.13	2.74	5.14	27.0 ^c	27.0 ^c	2.05	4.10
diaphragm life, d	300	275	200	200				
cell to cell distance, m	2.32	3.05	1.6	2.13			1.5	1.5
types of electrodes	monopolar	monopolar	monopolar	monopolar	bipolar	bipolar	monopolar	monopolar

^a All cells employ DSA anodes.

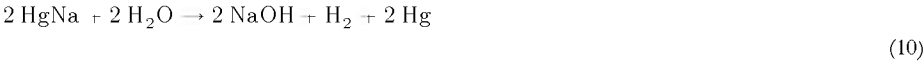
^b Designed to accommodate rectifier amperages from 30 to 150 kA. Cells of various kA ratings vary in the length of cell and number of anodes and cathodes. Operating characteristics vary with current density (30).

^c Capacity per electrolyzer.

Mercury Cells. The cathode material in mercury cells, mercury [7439-97-6], Hg, has a high hydrogen overvoltage. Hydrogen evolution is suppressed and sodium ion reduction produces sodium amalgam [11110-52-4], HgNa.



Because little hydrogen is generated at the cathode of mercury cells it is possible to construct cells without a diaphragm. Mercury (cathode) flows continuously down a long, gently sloping trough. The anodes, generally DSA, are held a minimum distance above the mercury. Chlorine is produced at the anode as in equation 1. Sodium amalgam leaving the mercury cell passes through a decomposer or denuder containing graphite where water is reduced to hydrogen and sodium is oxidized to caustic soda.



Mercury denuded of sodium is pumped back into the cell for continuous recirculation.

Caustic soda concentrations of 50% are produced directly from equation 11. This advantage is offset by higher operating cell voltages and some mercury contamination of the environment. This latter problem has been diminished or solved to an acceptable extent (31); however, it continues to influence the choice of cells for new plants. No new mercury cells have been installed in the United States since 1970 (32).

Membrane Cells. Membrane cells are separated into anode and cathode compartments by selective ion-exchange membranes. Sodium chloride brine is fed into the anode compartment. Chlorine is produced at the anode (eq. 1). Sodium ions and some water of hydration are transported through the membrane into the cathode compartment. Depleted brine is removed from the anode compartment saturated, treated, and returned to the cell. Hydrogen and 30–35% caustic soda are produced at the cathode. To achieve good caustic soda concentration control, some caustic soda is usually recirculated to the cathode compartment along with the required additions of water. The advent of selective ion-exchange membranes in 1971 made membrane cells practical (33). Membrane cells have many advantages over diaphragm cells and mercury cells, including higher caustic concentrations and nearly chloride-free caustic without mercury pollution. Mercury cells produce high quality caustic soda, but create a pollution problem.

Membrane cells are the state of the art chlor-alkali technology as of this writing. There are about 14 different membrane cell designs in use worldwide (34). The operating characteristics of some membrane cells are given in Table 3. The membranes are perfluorosulfonate polymers, perfluorocarboxylate polymers, and combinations of these polymers. Membranes are usually reinforced with a Teflon fabric. Many improvements have been made in membrane cell designs to accommodate membranes in recent years (35,36).

Table 3. Operating Characteristics of Membrane Cells

Parameter	Bipolar cells		Monopolar cells				
	Hoechst Uhde	Asahi Chemical Acilyzer	Chlorine Engineers DCM 406 × 2	ICI FM21 SP	Asahi Glass Azec	Oxytech MGC	Lurgi
current rating, kA	2–18	5.8–21.6	36–73	1–100	18–340	6–225	
current density, kA m ²	2–6	2–4	1.98–4	1.5–4.1	3–4	1.5–4.5	1–4
cell voltage, V	2.95–3.4	3.2–3.3	3.0–3.1		3.0–3.05		
energy (d-c) consumed, kWh t, NaOH	2080–2405	2150–2230	2000–2050	2150–2200		2050–2300	1850–2300
cells electrolyzer	2–120	50–100	12	1–120	30–540	2–30	2–10 ⁺
electrode area, m ²	0.9–2.7	2.7–5.4	3.03				~0.8
current efficiency, %	93–96	94–97			92–97		
height, mm			1300		1200	1960	1200
width, mm			2560		2400	1140	660
length, mm			1195			460–1900	

Chlorates. Sodium chlorate is produced by the electrolysis of sodium chloride at pH 6.5–7.5 in a one-compartment cell. DSA anodes and steel cathodes are generally used in chlorate cells. The electrolysis products, hypochlorous acid, and hypochlorite ions, react chemically to produce chlorate (eq. 4). Chlorate cells are designed with reaction volumes to provide time for equation 4 to proceed. Figures 2 and 3 show chlorate cells with their associated reaction volumes. The Chemetics cells in Figure 2 are separate from the reaction tank. Gas lift circulation causes the liquors to circulate through the cells into the reaction volume and back to the cells. The Huron cells in Figure 3 are submerged into the reaction tank. Liquors circulate through the cells by gas lift and by more dense cooled liquors descending from cooling coils.

Sodium dichromate [31924-34-2], Na₂Cr₂O₇, is added to chlorate cell electrolytes. The chromate minimizes the reduction of hypochlorite and chlorate at the cathode. Chromate also minimizes the corrosion of cathodes when or where cathodic current densities are inadequate to cathodically protect the steel cathodes from the very corrosive electrolyte. The reaction tanks are generally constructed of corrosion-resistant plastics. Huron cells (Fig. 3) are contained in a fiber-reinforced plastic tank with a proprietary lining. Chemetics cells (Fig. 2) are contained in titanium housings, and the reaction tank is titanium. The use of plastic materials in chlorate cells usually limits the operating temperature to about 65°C. Chlorate cell systems constructed predominately of metals allow higher operating temperatures, up to 80°C. Some chlorate plants operate several of the cell-reactor tank units in cascade fashion. Sodium chloride brine may enter one end of a series of Huron cell units, flow from one cell unit to the next until the cell liquor contains up to 600 g L of NaClO₃ and about 100 g L NaCl. Other plants operate each cell system at essentially final conditions. The entire electrolysis from NaCl to NaClO₃, cell liquor is done in a single-cell-reaction tank unit. Chlorate cell liquor is treated to convert as much hypochlorite to chlorate as is practical. Residual hypochlorite is destroyed to protect downstream equipment from corrosive attack. Chlorate liquors are cooled and evaporated in crystallizers to produce solid NaClO₃. Crystallizer liquors containing NaCl, NaClO₃, and chromate are resaturated with NaCl and recycled to the chlorate cells. Operating characteristics for some chlorate cells are given in Table 4.

Table 4. Operating Characteristics of Chlorate Cells

Parameter	Chemetics	Eka Nobel	Huron	Krebs Paris
cell type ^a	B	B	B	M (NC-24)
current, kA	22.5–87.5	25–36	20	77.4
current density, kA m ₂	1.5–3.5	1.0–1.2	3.2	2.4
current efficiency, %	95 ±	95–96	95 ±	94–96
energy (d-c), kWh t	4250 ±	4650–4800	5200	4910–5210
temperature, °C	80		65	80
pH	6.5	6.5–6.7	6.5–7.0	6.5–6.5
anode materials	DSA	DSA	DSA	DSA

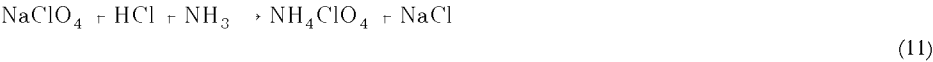
cathode material	Stahrm ^a et	steel	Ti	steel
product, g L				
NaClO ₃	650	550–600	600–650	540–580
NaCl	110	135	60–80	100–120
Na ₂ Cr ₂ O ₇ , g L	3	1.0	1.5	3.5–5
cells electrolyzer	up to 17	8	48	3

^a B bipolar; M monopolar electrodes.

Nearly 95% of the sodium chlorate produced in North America is used to produce chlorine dioxide [10049-04-4], ClO₂, for pulp (qv) bleaching (see BLEACHING AGENTS, PULP AND PAPER) (37). Minor amounts are used to produce other chemicals such as KClO₃, NaClO₂, NaClO₄, etc, to recover uranium [7440-61-1], U (see URANIUM AND URANIUM COMPOUNDS) and for agricultural uses as a defoliant or herbicide (see HERBICIDES).

Perchlorates. Concentrated solutions of sodium chlorate are electrolyzed to produce sodium perchlorate [7601-89-0], NaClO₄ (see PERCHLORIC ACID AND PERCHLORATES). The electrolytic cells for perchlorate production are usually simple tank electrolyzers. Anode materials are limited to platinum, platinum-clad titanium, or lead dioxide on graphite or titanium. The cathodes are generally mild steel. Cell feed solutions may contain 600 g L of NaClO₃. Production of very high (1000 g L or more) concentrations of NaClO₄ is achieved by resaturating partially converted electrolyte with NaClO₃ and continuing the electrolysis to minimum final NaClO₃ concentrations. A mechanism of perchlorate formation has been proposed involving the reaction of chlorate ion and an absorbed active oxygen species at the anode (38,39).

Most of the NaClO₄ produced is converted to ammonium perchlorate [7790-98-9], NH₄ClO₄.



The products of equation 11 are separated by controlled crystallizations to produce high purity crystalline anhydrous ammonium perchlorate and sodium chloride. The main use for ammonium perchlorate is as an oxidizer in the propellant of rockets and missiles (see EXPLOSIVES AND PROPELLANTS).

Manganese Dioxide. High performance alkaline batteries (qv) and some lithium batteries require pure electrolytic manganese dioxide [1313-13-9] (EMD), MnO₂. Other forms of MnO₂, natural ore (NMD) and chemical manganese dioxide (CMD), are used in large batteries where performance requirements are less demanding. The production of EMD involves the following unit operations: first, high quality manganese ores are reduced. Reduced ore is then leached using acidic electrolyte from the cells and makeup sulfuric acid [7664-93-9], H₂SO₄. Crude manganese sulfate [10124-55-7], MnSO₄, solution is treated to remove impurities. Purified manganese sulfate solution is fed into an electrolytic cell where MnO₂ is deposited on the anode. Deposited MnO₂ is periodically harvested, milled, neutralized, and packaged for market (see MANGANESE COMPOUNDS).

Electrolytic cells for EMD production consist of open-top tanks having titanium anodes and graphite or metallic cathodes. No separators are required. High quality EMD is produced at very low (60–80 A m²) anode current densities on titanium anodes. Suspensions of small particles of EMD in the electrolyte of EMD cells allow good quality EMD to be produced on titanium anodes at higher current densities, 100–200 A m². The use of suspensions in EMD cell electrolytes was first patented by Japan Metals and Chemicals Co., Ltd. in 1976 (40). Numerous other publications and patents have appeared (41–44). A novel fiber glass reinforced plastic (FRP) isotenoid cell tank design is being used in Australia. The tank consists of a suspended FRP bag supported in a rectangular steel frame (45). Table 5 details some operating conditions for EMD production.

Table 5. Production Conditions for Electrolytic Manganese Dioxide

Parameter	Typical ^a	Suspensions ^b
MnSO ₄ feed solution, g L	130–160	105 ^c
electrolyte		
H ₂ SO ₄ , g kg	30 ± 5	30–50 ^d
MnSO ₄ , g kg	40 ± 10	^e
anode	Ti	Ti
cathode	graphite	graphite
temperature, °C	95–98	90–95
cell voltage, V	2.7–3.0	2.4–3.2
current density, A m ²	60–80	85–150

^a Ref. 46.

^b Ref. 42.

^c NaOH is added to a pH of 7.8⁺. Air agitation is employed.

^d Value is g L.

^e MnSO₄ plus 10–200 mg L of hydrated manganese oxide.

The electrolytic deposition of MnO₂ at an anode may be described by



The cathode reaction is the reduction of water to produce hydrogen (eq. 8). The mechanism of the anodic deposition of MnO₂ has been investigated (44,47).

World production capacity for electrolytic manganese dioxide was estimated to be about 198,000 t yr in 1988 (48). An EMD production facility having 18,000 t yr capacity was built in Australia by Australian Manganese Co., Ltd., in 1987–1989 (46). U.S. production in 1991 was about 39,000 t (7).

Water Electrolysis. The electrolysis of water for hydrogen and oxygen production is economically attractive only in those areas where electric power is available at very low cost. Hydrogen (qv) is usually the primary product (see HYDROGEN ENERGY); oxygen (qv) is a co-product. Hydrogen in large quantities is produced by steam (qv) reforming methane, ie, natural gas (see GAS, NATURAL). Hydrogen produced in chlor-alkali and chlorate cells is available in some areas, and in some cases is compressed and sold in cylinders and as liquid hydrogen. Oxygen is commonly produced by the separation of air (see

Cryogenics).

Total production of hydrogen in the United States in 1988 was $61.5 \times 10^9 \text{ m}^3$ (49). Total hydrogen production by electrolysis of water in 1988 was about $5.7 \times 10^6 \text{ m}^3$ (<0.01%). U.S. capacity for liquid hydrogen production as of June 1989 was about 42,500 t yr, equivalent to about $481 \times 10^6 \text{ m}^3$ yr.

Cells for the electrolysis of water are available from several sources. These cells have been described (50,51). Water electrolysis cells must operate at low voltages to achieve good energy efficiency. The theoretical decomposition voltage for hydrogen and oxygen production is 1.23 V. Actual cell voltages are 1.8–2.6 V. Current efficiencies closely approach 100%. Cells are usually of a filter-press design incorporating bipolar electrodes, porous diaphragms or ion-exchange membranes, alkaline electrolyte, KOH, and catalyzed electrodes. Most cells operate at high pressures, about 3 MPa (30 atm) (52,53).

Recent developments in water electrolysis cells include the use of ion-exchange membranes (54), optimization of pressure and temperature (55), and suggestions for better electrolysis cell designs (56). Interest in water electrolysis is associated with load leveling and space exploration (57). Electric power available during periods of low usage, ie, off-peak periods, can be used to electrolyze water. The hydrogen and oxygen produced can be stored and used in fuel cells (qv) to produce electric power when peak power demands occur.

Heavy water [11105-15-0], D₂O, was produced by a combination of electrolysis and catalytic exchange reactions. Some nuclear reactors (qv) require heavy water as a moderator of neutrons. Plants for the production of heavy water were built by the U.S. government during World War II. These plants, located at Trail, British Columbia, Morgantown, West Virginia, and Savannah River, South Carolina, have been shut down except for a portion of the Savannah River plant, which produces heavy water by a three-stage process (see Deuterium and Tritium): an H₂S–H₂O exchange process produces 15% D₂O; a vacuum distillation increases the concentration to 90% D₂O; an electrolysis system produces 99.75% D₂O (58).

Cold fusion has been reported to result from electrolyzing heavy water using palladium [7440-05-3], Pd, cathodes (59,60). Experimental verification of the significant excess heat output and various nuclear products are still under active investigation (61,62) (see Fusion Energy).

Fluorine. Fluorine is the most reactive product of all electrochemical processes (63). It was first prepared in 1886, but important quantities of fluorine were not produced until the early 1940s. Fluorine was required for the production of uranium hexafluoride [7783-81-5], UF₆, necessary for the enrichment of ²³⁵U (see Diffusion Separation Methods). The Manhattan Project in the United States and the Tube Alloy project in England contained parallel developments of electrolytic cells for fluorine production (63). The principal use of fluorine continues to be the production of UF₄ from UF₆.

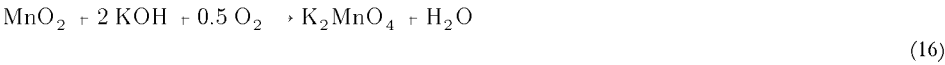


The electrolyte used in fluorine cells is KF–HF in a ratio that minimizes melting point, HF vapor pressure, and corrosion of materials. Various ratios have been used. The manufacture of fluorine in the early 1990s was based on the electrolysis of KF:2HF, which allows cell operating temperatures of 100–105°C.

Fluorine has been compressed, liquified, and shipped. However, most fluorine is produced and used on site. Fluorine production in the United States is based on electrolytic cells developed in the 1940s. Modern type "E" cells are rated for 6 kA (64).

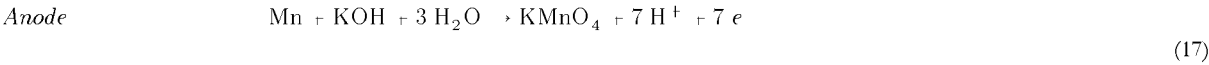
Fluorine cells use carbon anodes, steel cathodes, and nickel or monel wire mesh diaphragms contained in a monel tank with a water jacket for cooling. Cells operate at about 6 kA, 10–12 V, and 95–105°C. Energy consumption for fluorine production is about 22 kWh/kg. Figure 6 is a diagram of a fluorine cell. Fluorine cell voltages are increased by the formation of fluorinated carbon. Conditions for the formation of (CF)_n and (C₂F)_n on anode surfaces have been studied (65). Cell voltages have been decreased somewhat by impregnating carbon anodes with molten fluoride at high temperatures and pressures (66).

Permanganate. Potassium permanganate [7722-64-7], KMnO₄, is produced in commercial quantities. It may be prepared from manganese dioxide or directly from manganese metal, ferromanganese, or other manganese alloys (67). The Carus Chemical Co. produces potassium permanganate from manganese dioxide in the United States. First potassium manganate [10294-64-1], K₂MnO₄, is prepared by a liquid-phase oxidation of manganese dioxide with potassium hydroxide in air



The reaction mixture is filtered. The solids containing K₂MnO₄ are leached, filtered, and the filtrate composition adjusted for electrolysis. The solids are gangue. The Carus Chemical Co. electrolyzes a solution containing 120–150 g/L KOH and 50–60 g/L K₂MnO₄. The cells are bipolar (68). The anode side is monel and the cathode mild steel. The cathode consists of small protrusions from the bipolar unit. The base of the cathode is coated with a corrosion-resistant plastic such that the ratio of active cathode area to anode area is about 1 to 140. Cells operate at 1.2–1.4 kA. Anode and cathode current densities are about 85–100 A/m² and 13–15 kA/m², respectively. The small cathode areas and large anode areas are used to minimize the reduction of permanganate at the cathode (69). Potassium permanganate is continuously crystallized from cell liquors. The caustic mother liquors are evaporated and returned to the cell feed preparation system.

The direct electrochemical oxidation of manganese alloys was developed and commercialized at the Rustavi Chemical Combine in the Georgian Republic (formerly the USSR). The electrode reactions are

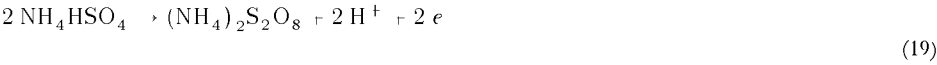


The anodes used were cast ferromanganese; the electrolyte, KOH–K₂CO₃, and current efficiencies for this process were about 40%. Energy requirements for this process, about 15 kWh/kg of KMnO₄, plus the cooling requirement to maintain cells at 20°C, made this process uneconomical (70,71).

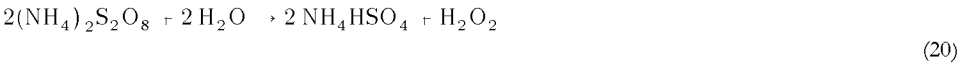
Production of potassium permanganate in the CIS is believed to be from potassium manganate. Electrolysis of potassium manganate in a continuous-flow electrolytic cell with turbulent electrolyte flow and continuous crystallization has been reported (72).

Direct electrochemical oxidation of manganese dioxide suspended in KOH to produce potassium permanganate has been patented (73). It is not known whether this is a commercial process. Metal cathodes having adherent surface deposits of manganese dioxide improve permanganate production efficiencies by decreasing the manganate and permanganate reduction reactions at the cathode (74). Production capacity for potassium permanganate in the United States was about 21,000 t/yr in 1991 (8).

Hydrogen Peroxide. Peroxydisulfuric acid [13445-49-3], H₂S₂O₈, is one of the strongest oxidizing agents known. It and other peroxydisulfates are produced electrochemically (75). The production of peroxydisulfates was once important for the manufacture of hydrogen peroxide [7722-84-1] (qv), H₂O₂ (76). In 1956 more than 80% of the production of H₂O₂ was from peroxydisulfates. Peroxydisulfates are produced by electrolyzing bisulfates or sulfates in acidic solution at smooth platinum anodes:



Hydrogen peroxide was produced by hydrolysis:



The development of the autoxidation of alkyl anthraquinones led to a rapid increase in the production of H₂O₂ but a sharp decline in the importance of the electrolytic process. In 1991 the total North American, Western European, and Japanese capacity for H₂O₂ production was more than 870,000 t (77). No H₂O₂ was produced by the electrolytic peroxydisulfate process. The last plant using this process closed in 1983.

Rapid growth in demand for H₂O₂ in the late 1980s created widespread interest in new processes for H₂O₂ production (78). New process development efforts were generally focused either on direct combination of hydrogen and oxygen to produce H₂O₂ (79) or the electrolytic reduction of oxygen to produce H₂O₂. HD Tech Inc., a joint venture of Dow Chemical Canada Inc. and Huron Technologies Inc., has developed and commercialized an electrolytic reduction of oxygen process (77,78). The electrolytic cell is divided into two compartments by a porous separator. A solution of caustic soda and sodium silicate is fed into the anode compartment. Oxygen is produced at the anode. Oxygen enters the top of a catalytic trickle bed cathode and is reduced to form H₂O₂ and caustic soda. The alkaline peroxide product has a H₂O₂ to NaOH ratio of about 1 to 1.7, and can be used immediately for bleaching pulp (80). This process is best suited for plants that use five or more tons per day of H₂O₂, and are 500 km or more from a commercial H₂O₂ production source.

Electrowinning of Metals

The metals that are produced by electrolysis (81) are included in Table 1. Fused salt processes are used when the reactivity of the metal does not allow electrowinning from aqueous solutions. Manganese is the most reactive metal that is produced by electrolysis of an aqueous solution.

Electrowinning from Aqueous Solutions. The aqueous processes for electrowinning of metals from ores have the following common unit operations or steps: (1) the metal in the ore is converted to an acid-soluble form and this may be an oxidizing roast or a reduction; (2) ores from step 1 are leached, usually in sulfuric acid; (3) metal solutions from step 2 are purified and in some cases concentrated; (4) purified metal solutions are electrolyzed in cells where the metal is deposited on the cathode; and (5) acid is produced at the anode and recycled to the leaching step 2. Some acid values are lost, usually in the purification step, 3. Makeup acid is added in the leaching step, 2. In most cases the metal solution from leaching step 2 contains impurities, other metals. Many of these metals have the characteristic of low hydrogen overvoltage. Codeposition of the impurity metals causes contamination of the desired product and decreases current efficiencies. The removal of impurities before electrolysis is very important. This is especially true in the case of the more reactive metals such as zinc [7440-66-6], Zn, and manganese [7439-96-5], Mn. These metals have deposition potentials close to the hydrogen evolution potential. The current efficiency of manganese electrowinning is about 60 to 68° o. The principal inefficiency is hydrogen evolution.

The electrowinning of metals from aqueous solutions is generally carried out in tank cells. Figure 7 shows a typical metal electrowinning cell. Process conditions for electrowinning metals from aqueous solutions are given in Table 6 (82,83). Developments in the electrowinning of metals from aqueous solutions have been directed toward improved anodes, improved additives, higher current densities, the use of ion-exchange membranes, better electrolyte quality control, and computer modeling of the processes.

Table 6. Production of Metals by Electrolysis of Aqueous Solutions^a

Metal	Anode	Diaphragm	Cathode	Cell feed ^b , g L	Electrolyte ^b , g L	Temperat ure, °C	Cell voltage, V	Cathode current density, A m ²	Energy requirement, kWh kg	Current efficiency, ° o
Cd	Pb–Ag	no	Al	90–200 Cd, 20–40 Zn	10–20 Cd, 20–40 Zn, 60–140 H ⁺ ^c	30–35	2.5–2.7	80	1.5	90
Cr	Pb–Ag	yes	316 stainless steel	20–70 Cu, 20–70 H ⁺	^d	53	4.2	700	18	45
Cu	Pb–Sb–Ag	no	Cu	20–70 Cu, 20–70 H ⁺	^d	30–35	2.0–2.2	130	2.2	80–90
Mn	Pb–Ag	yes	stainless steel or Ti	30–40 Mn, 125–150 NH ₄ ⁺ , 0.1 SO ₂ + glue ^d	^d	35	5.1	400–600	8–9	60–68
Ni	Pb	yes	Ni 99.9° o		^e	52	3.4	180		91–96
Zn	Pb–Ag	no	Al	100–200 Zn	100–200 H ⁺ , 20–40 Zn	35	3.2–3.6	350–1000	3.3	90

^a Refs. 82,83.
^b As sulfates unless otherwise noted.
^c Anolyte in g L: 13 Cr(VI), 2 Cr(III), 24 NH₃, 280 H₂SO₄. Catholyte in g L: 11.5 Cr(III), 12.5 Cr(II), 84 NH₃; pH 2.1–2.4.
^d Anolyte in g L: 10–20 Mn, 25–40 H₂SO₄, 125–150 NH₄⁺. Catholyte in g L: pH 6–7.2.
^e Catholyte in g L: 70 Ni(II) + H₃BO₃, Na₂SO₄; pH = 3.0–3.5. Anolyte in g L: 40 H₂SO₄.

Corrosion rates of lead–silver anodes containing various amounts of silver were studied under various operating conditions. Data are available for optimizing anode composition for minimum overall cost (84). Titanium anodes coated with various catalytic materials have been developed for metal electrowinning cells (85–87). Hydrogen diffusion anodes are being developed (88). These anodes would consume hydrogen in fuel cell fashion and decrease cell voltages by about 1.8–2.0 V.

Additives are used in metal electrowinning to improve the quality of deposited metal, increase current efficiencies, and decrease the work required to remove metals from cathodes. Improvements in these additives have been developed (89,90).

Production rates and profits can generally be increased when higher current densities can be applied to electrowinning cells. Capital investments required to achieve higher current densities are a critical factor. Zinc electrowinning on rotary drum cathodes produces continuous ribbons of zinc at very high current densities, up to 5000 A m² (91,92). Small-scale cells have been operated, but no commercial applications are known as of this writing (93). Ion-exchange membranes may be used in some metal electrowinning cells (94,95).

Systems for evaluating electrolytes for metal electrowinning have been developed and are being used commercially in zinc production (96). Computerized mathematical models of zinc electrowinning cells have been developed and validated by comparison with experimental data taken from pilot-plant cells (97).

Electrowinning from Fused Salts.

Aluminum. Aluminum [7429-90-5], Al, is produced worldwide by the Bayer-Hall-Heroult process. This process involves the electrolysis of alumina [1344-28-1], Al₂O₃, dissolved in molten cryolite [15096-52-3], Na₃AlF₆ (see ALUMINUM AND ALUMINUM ALLOYS). The electrolytic cells or pots operate at about 975°C. The overall reaction is



Graphite anodes are oxidized to CO₂ and CO. Cell gas contains about 70–90% CO₂. Reactions occurring in aluminum cell cathodes have been studied extensively (98). The cells are constructed from large steel tanks lined with refractory insulating bricks, then lined with an inner covering of graphite which conducts current to the molten aluminum cathode pool. The graphite anodes are of two types, prebaked anodes and Soderberg anodes. The prebaked anodes are large graphite blocks that must be replaced as they are consumed in equation 21. The Soderberg anode is continuously formed by addition of a carbon pitch paste to a mold held above the molten electrolyte in the cell. The paste is added at the rate of graphite consumption in the cell. The whole mass moves slowly down toward the melt and is baked *in situ* (see CARBON).

Developments have taken place in the areas of improved insulation, pot linings, and cathode and anode materials. A compressible insulation was developed to minimize potlining heaving (99). Potlining heaving is caused by sodium in the graphite linings. Substantial increases in pot lives are to be expected when new graphite cathodes are preheated using gas burners. Cathode surfaces are heated slowly up to about 700°C (100). Molten pools of cathodic aluminum have rippling surfaces. This is the result of magnetohydrodynamic effects. Large anode to cathode gaps are required to prevent contact of molten aluminum to the anodes. This results in high voltages. Aluminum cells with drained cathodes, that are molten aluminum wettable, allow smaller electrode gaps and lower cell voltages. Coatings of titanium diboride [12045-63-5], TiB₂, and titanium carbide [12070-08-5], TiC, on graphite have been used in these cells (101). The development of anode materials for aluminum cells has met with limited success (102,103). The electrical energy required to produce aluminum has been decreased from about 55 d-c kWh/kg in 1900 to about 16 d-c kWh/kg in 1980 (104). An excellent discussion of energy consumption and efficiency of aluminum production and other electrochemical processes is found in Reference 104.

Sodium. Sodium is produced by the electrolysis of a fused salt mixture of calcium chloride [10043-52-4], CaCl₂, and NaCl in a Downs cell (105). CaCl₂ is used to lower the melting point of the electrolyte. Operating temperatures in the 570 to 590°C range are practical using NaCl–CaCl₂ mixtures. The Downs cell consist of several graphite anodes surrounded by steel cathodes. A metal diaphragm is placed between each anode and cathode to prevent contact of chlorine and sodium. The respective streams rise to the top of the cell. Chlorine is collected in a hood, and the liquid sodium and calcium are channeled to a riser pipe. The collected metal is allowed to cool to 110°C to precipitate the calcium [7440-70-2], Ca, metal codeposited with sodium. The calcium is filtered out and the liquid sodium contains less than 0.04 wt % calcium (see SODIUM AND SODIUM ALLOYS). Eutectic melts of BaCl₂, CaCl₂, and NaCl have been studied by electrochemical means (106). An improved cell for the production of sodium was patented (107). This cell utilizes an electrode separator, solid electrolyte tubes that are permeable to the flow of sodium ions but impermeable to fluids and the flow of other ions. Production of sodium from mixtures of NaCl and aluminum chloride [7446-70-0], AlCl₃, has been described (107).

Lithium. Several processes for lithium [7439-93-2], Li, metal production have been developed. The Downs cell with LiCl–KCl electrolyte produces lithium in much the same manner as sodium is produced. Lithium metal or lithium–aluminum alloy can be produced from a mixture of fused chloride salts (108). Granular Li metal has been produced electrochemically from lithium salts in organic solvents (109) (see LITHIUM AND LITHIUM COMPOUNDS).

Magnesium. There are three electrolytic processes for magnesium [7439-95-4], Mg, production: the Dow process, a process developed by I.G. Farbenindustrie in Germany, and an Alcan process. All processes involve the electrolysis of magnesium chloride [7786-30-3], MgCl₂. The Dow process uses partially dehydrated cell feed, and the German process requires completely anhydrous MgCl₂ as cell feed (110). The Dow cell is a large steel pot about 1.5 m wide, 1.8 m deep, and 4.0 m long. Graphite anodes are suspended through an arched refractory cell cover. The steel pot and steel baffles around each anode are the cathode. Chlorine is collected from the arched cell cover of the Dow cell. Magnesium rises to the top of the electrolyte and is directed through inverted troughs to a storage well from which it is removed.

The Alcan process has been used commercially by Osaka Titanium Co. in Amagasaki, Japan. Multipolar cells of 1000 t/yr capacity are in operation. Energy consumption is about 9.5–10 kWh/kg of magnesium metal (111).

Research and development efforts have been directed toward improved cell designs, theoretical electrochemical studies of magnesium cells, and improved cathode conditions. A stacked-type bipolar electrode cell has been operated on a lab scale (112). Electrochemical studies of the mechanism of magnesium ion reduction have determined that it is a two-electron reversible process that is mass-transfer controlled (113). A review of magnesium production is found in Reference 114.

Beryllium. Beryllium [7440-41-7], Be, metal is produced by electrolysis of KCl–NaCl–BeCl₂ melts. Temperatures up to 900°C are required. Cell voltages are 6 to 9 V (115). Electrolysis of mixtures of beryllium oxide [1304-56-9], BeO, in lithium fluoride [7789-24-4], LiF, and beryllium fluoride [7787-49-7], BeF₂, has produced beryllium metal at about 700°C and 2.6 V (116). Details of fused salt metal winning processes are given in Table 7.

Table 7. Production of Metals by Electrolysis of Fused Salts

Metal process	Anode	Cathode	Electrolyte melt, %	Temperature, °C	Cell voltage, V	Cathode current density, kA/m ²	energy (d-c) consumption, kWh/kg	Current efficiency, %
Al Hall-Heroult	prebaked graphite or Soderberg graphite	Al	80–85 cryolite, 5–7 CaF ₂ , 5–7 AlF ₃ , 2–8 Al ₂ O ₃	940–980	4.1	3–3.5	13–15	92–95
Na Downs	graphite	steel	58 CaCl ₂ 42 NaCl	580 ± 10	5.7–7.0	9.7	10–11	80–85
Li Downs	graphite	steel	LiCl–KCl eutectic	450	6.8	5	36	
Mg DOW	graphite	steel	20 MgCl ₂ 20 CaCl ₂ 60 NaCl	700–720	6.3		18.5	75–80

Electrochemical Waste Treatment

In many instances the metal processing industry produces aqueous effluents containing dissolved metals. Many of these metals are toxic and controlled by Environmental Protection Agency (EPA) regulations. Most water treatment plants do not remove toxic metals, or they concentrate the toxic metals in sludges that are classified as hazardous waste. Compliance with EPA regulations adds to overall production costs, and adversely affects the competitive viability of some plants. Therefore, the capital and operating costs for waste treatment systems should be as low as possible. Concentrations of the metals in wastewater are generally very low. The flow rates may be relatively high. Electrodes with large areas operating at low current densities are desirable for this application. Various porous electrodes, particulate-bed electrodes, fluidized-bed electrodes, and roll cells (117) have been developed for metals recovery from dilute wastewater streams. Electrochemical processing has advantages over other chemical processes (117) because the electrochemical process usually requires no addition of materials. Chemical processes require the addition of chemicals that usually expand the volume and complexity of the overall waste disposal process. Electrochemical processes for wastewater treatment often provide recovery of metal resources that partially offset processing costs (118,119). Silver has been recovered electrolytically from spent photographic liquors for many years (120) (see PHOTOGRAPHY). Silver recovery was motivated by economics. More recently, heavy-metal recovery has been motivated by environmental regulations (see HAZARDOUS WASTE TREATMENT; RECYCLING; SILVER AND SILVER ALLOYS).

In addition to metal recovery, which generally involves cathodic reduction, some waste may be treated by anodic oxidation processes. Many organic contaminants in wastewater can be oxidized by electrochemical treatments. High overvoltage anodes have been developed to improve the efficiency of these oxidations (121). Electrochemical oxidation of halogenated hydrocarbons has been achieved using barium peroxide [1304-29-6], BaO₂, in aqueous NaCl solutions. Cationic surfactants are used to suspend insoluble organic compounds and the BaO₂ (122). Carbon tetrachloride [56-23-5], CCl₄, has been oxidized by this system (123). Several cell systems are commercially available for wastewater treatment. One such cell system is that offered by Eltech International Corp. shown in Figure 9.

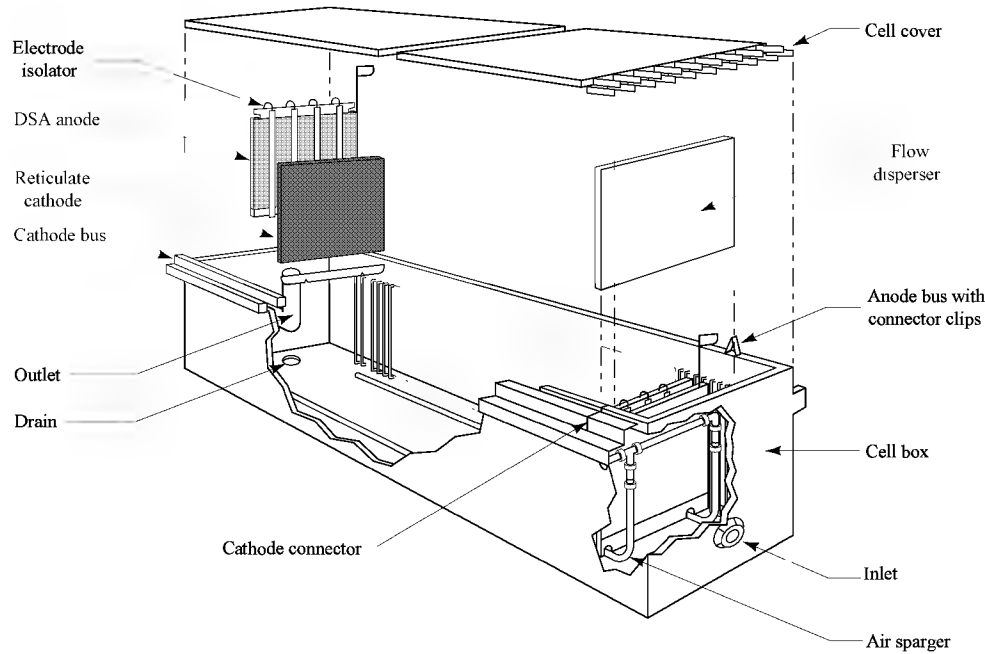


Fig. 9. Retec cell for heavy-metal recovery from wastewaters.

Courtesy of Eltech International Corp.

Electrolytically generated hypochlorite may be used for the oxidative destruction of cyanides (qv) or the sterilization of domestic wastes. Several on-site systems for swimming pool sterilization and municipal waste treatment works have been developed. One of these systems is described in Reference 124. On-site production and immediate use of chlorine is considered safer than the transportation of chlorine.

Some examples of electrochemical waste treatment are given in Table 8 (122,123,125–132). Other electrochemical processes such as electrodialysis (qv), electroflotation, and electrodecantation are also used in waste treatment.

Table 8. Electrochemical Treatment of Wastewater

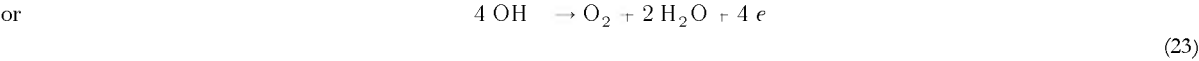
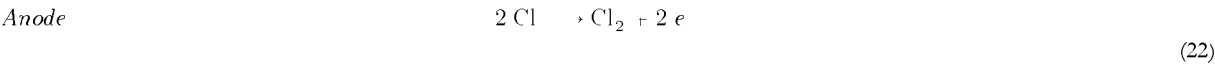
Contaminant	Reaction	Source	Electrochemical system for treatment	References
Ag	$\text{Ag}^+ + e^- \rightarrow \text{Ag}$	photographic fixers, etc	Eltech Retec cell	126,127
			Du Pont ESE cell	128
			Swiss roll cell	129
			rotating cylinder cell	125
Cu	$\text{Cu}^{+} + 2 e^- \rightarrow \text{Cu}$	Cu plating rinse waters, etc	Swiss roll	129
			Du Pont ESE	128
			rotating cylinder	125
			fixed-bed flow-through anode	130
Fe	$\text{Fe}^{2+} \rightarrow \text{Fe}^{3+} + e^-$	acidic coal mine waters	Eltech Retec chrome recovery cell	131
Cr	recovery of CrO ₃	Cr plating rinse water	Swiss roll	129
Cr	$\text{Cr(VI)} + 3 e^- \rightarrow \text{Cr(III)}$	Cr plating wastewater	plate and frame filter press cell, SnO ₂	121
organic compounds	oxidation	chemical process water	anodes	
phenol in water	oxidation	process wastewater	Pt anodes	132
halogenated	destructive oxidation	chemical process water	BaO ₂ suspensions in aqueous	123
hydrocarbons			NaCl plus cationic surfactants at graphite anodes	122

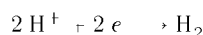
Safety and Environmental Considerations

The electrochemical process industries are confronted with a wide range of hazards. These include electrical hazards, various explosion hazards, and the hazards associated with exposure to reactive chemicals.

Electrical Hazards. Electrical hazards include all those associated with both alternating and direct currents. Cell components carrying direct current are commonly exposed in cell rooms and tank houses. Most cell circuits are insulated from ground and designed to minimize potential difference from ground potential. A direct short to ground in an otherwise ungrounded circuit causes little if any damage. Strong magnetic fields exist in cell rooms. Safety rules generally prohibit magnetic and conductive materials beyond a specified length in the cell rooms. This is to avoid accidental contact with exposed electrical conductors.

Explosion Hazards. The electrolysis of aqueous solutions often lead to the formation of gaseous products at both the anode and cathode. Examples are hydrogen and chlorine from electrolysis of NaCl solutions and hydrogen and oxygen from electrolysis of water. The electrode reactions,



Cathode

(24)

are separated by a diaphragm or membrane to obtain pure products. Mixing of these gases is potentially dangerous because the range of explosive mixtures is very wide. This is clearly seen in Figure 10, which shows the explosion limits for chlorine–hydrogen–oxygen mixtures (133).

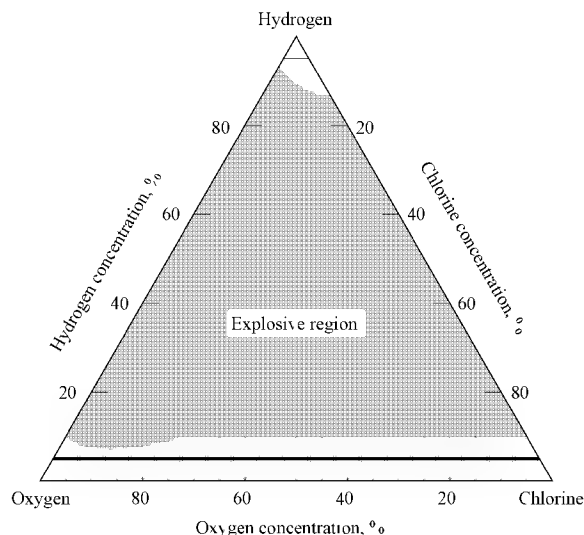


Fig. 10. Hydrogen–chlorine–oxygen explosive limits.

In chlor-alkali production the events that lead to a dangerous mixture of gases are dependent on the cell type. The following problems can lead to H_2 – Cl_2 mixtures in diaphragm cells. Hydrogen may diffuse through thin spots in diaphragms. Hydrogen may escape through the diaphragms when anolyte levels fall below the top of diaphragms. Excess back pressure from the hydrogen headers may force H_2 through the diaphragm. Metallic impurities in feed brine may precipitate in the alkaline diaphragm, become reduced to metals, and produce H_2 on the surface of the diaphragm. The normal operation of diaphragm cells produces chlorine containing 0.3 to 0.5% H_2 . The drying, compression, and liquefaction of this chlorine produces an explosive mixture of Cl_2 and H_2 if dilution gases are not introduced into the system.

Chlorine–hydrogen hazards associated with mercury cells result from mercury pump failures; heavy-metal impurities, particularly those with very low hydrogen overvoltage, ie, Mo, Cr, W, Ni; excessively low pH of feed brine; low NaCl concentrations in feed brine; and poor decomposer operation, which leads to high sodium amalgam concentrations in the cell.

Membrane cells generally produce high quality chlorine. Higher than normal H_2 concentrations in Cl_2 indicate that holes exist in the membrane.

Modern cells for sodium chlorate production are diaphragmless. Hydrogen produced at the cathode is recovered and used as fuel or for higher value uses. The composition of gases from the cells is normally about 97.5% H_2 , 2% O_2 , and 0.5% Cl_2 . These cells must operate at high current efficiencies to avoid dangerous H_2 – Cl_2 – O_2 mixtures. Conditions that cause explosive mixtures in these cells include low chloride concentrations, inadequate chromate concentrations, impurities that catalyze the decomposition of hypochlorite to chloride and oxygen, and loss of electric power. Gases released from chlorate cell liquors are particularly dangerous. Tanks for these liquors are either open-top tanks or tanks that are purged with air or nitrogen. Liquors exiting chlorate cells contain dissolved H_2 , very small H_2 bubbles, and hypochlorite. The escape of H_2 and the decomposition of hypochlorite to produce O_2 can produce explosive mixtures above chlorate liquors. Most modern cell plants have automatic air or nitrogen purges on cells and other critical equipment. Instrumentation usually includes continuous on-stream gas analyzers, alarms, and automatic purge systems.

When hydrogen recovery is unsafe or otherwise impractical it is vented into the cell room or tank house. The light weight and high diffusion rate of hydrogen and good cell room ventilation have made this an accepted practice.

Explosions have occurred in aluminum and magnesium plants as a result of molten metals coming into contact with water (134).

Reactive Chemicals Exposure. The hazards associated with exposure to reactive chemicals vary with the chemicals produced. The multitude of chemicals produced electrochemically precludes a detailed discussion of these hazards. Material Safety Data Sheets (MSDS) are available from suppliers. Anyone contemplating the production of a chemical must become thoroughly familiar with all safety, health, and environmental aspects of such production.

Technical and trade organizations are concerned with safety and the environment. The Chlorine Institute in North America and Euro Chlor in Western Europe are examples of organizations dedicated to the safe production, transport, and use of chlorine. Hazard and operability studies (HAZOP) reviews for new designs, plants, and expansions (135) have become required by policy in many operating companies. Papers on safety and environmental subjects are given at most technical meetings (136–138).

Access to cell rooms is usually restricted to authorized personnel. Required personal safety equipment includes rubber shoes or overshoes, goggles, hard hat, and some sort of emergency respirator. Safety training courses are generally required before one is allowed to enter a cell house without a guide.

The Chemical Manufacturers Association (CMA) and its members are also concerned about health, safety, and environmental problems. Responsible Care (139), a program initiated by the CMA in 1988, is designed to make day-by-day improvements in all aspects of health, safety, and the environment.

Although the electrolytic process industries are confronted with a wide range of hazards, their safety record has been excellent. The U.S. Bureau of Labor has reported lost work days (injuries and illnesses) per 100 full-time workers for the year 1990 (140).

Economic Aspects

Some electrochemicals are produced in very large quantity. Chlorine and sodium hydroxide production in 1991 were 10,727,000 t and 11,091,000 t, respectively (1). Aluminum was produced at the rate of 4,100,000 t yr and had an annual market value of about \$5.4 billion. Other electrochemically produced products are required in smaller volume. The production of the metals cadmium, lithium, and nickel were at the rates of 1600 t, 2800 t, and 8400 t, respectively for 1991 (see Table 1). Electrochemical processing plants produce a variety of products in a wide range of capacities.

Most large electrochemical processing facilities are located where raw materials, including electric power, are readily available at reasonable costs. Other factors influencing the location of electrochemical plants are proximity to markets, established transportation facilities, availability of water, and a source of labor. Large electrochemical plants are capital intensive, requiring large capital investment cost per employee.

The total annual value of products produced by electrochemical processing (see Table 1) is roughly \$18 billion. Electric power production in 1990 was 2908×10^9 kWh (9). Production of electric power increased about 2% in 1991 (4) although the economy was in a recession; thus, electric power production in 1991 was about 2966×10^9 kWh. About 5% of the total electric power produced is used by the electrochemical process industries. Published

industrial rates for electric power ranged from a low of 4.15 ¢ kWh from Public Service of Oklahoma to a high of 11.1 ¢ kWh from Long Island Lighting (4). Many of the largest plants have long-term electric power contracts at considerably lower power cost. Electric power cost is the largest single cost item in producing most large-volume electrochemicals.

BIBLIOGRAPHY

"Electrochemistry" in *ECT* 1st ed., Vol. 5, pp. 495–549, by G. C. Akerlog, Mellon Institute of Industrial Research; in *ECT* 2nd ed., Vol. 7, pp. 784–841, by G. C. Akerlog, Princeton University; "Inorganic" under "Electrochemical Processing" in *ECT* 3rd ed., Vol. 8, pp. 671–695, by P. M. Robertson, Eidgenössische Technische Hochschule.

1. *Chem. Eng. News*, 17, (Apr. 13, 1992).
2. *Chem. Mark. Rep.*, 30–38 (Oct. 5, 1992).
3. *Mineral Commodity Summaries* 1992, U.S. Dept. of Interior, Bureau of Mines, Washington, D.C., Jan. 1992, pp. 22–23.
4. J. B. Talbot and S. D. Fritts, *J. Electrochem. Soc.* **139**, 2981–3018 (1992).
5. *Chem. Mark. Rep.*, 42, (June 8, 1992).
6. *Chem. Mark. Rep.*, 42, (June 1, 1992).
7. Unpublished information from Kerr-McGee Chemical Corp., Oklahoma City, Okla., June 1, 1992.
8. R. Will, *Manganese, Chemical Economic Handbook*, SRI International, Menlo Park, Calif., June 1991, p. 233.4000P.
9. H. S. Burney and J. B. Talbot, *J. Electrochem. Soc.* **138**, 3156 (1991).
10. J. Dykstra and co-workers, *Ind. Eng. Chem.* **47**, 883 (1955).
11. J. O'M. Bockris and co-workers, eds., *Comprehensive Treatise of Electrochemistry*, Vol. 2, Plenum Press, New York, 1981, pp. 3–9.
12. R. E. White, ed., *Electrochemical Cell Design*, Plenum Press, New York, 1984, pp. 147–150 and 311–333.
13. R. J. Allen and co-workers, *St. Louis Meeting of the Electrochemical Society*, abstract #399, St. Louis, Mo., May 1992.
14. V. de Nora and J. W. Kuhn Von Burgsdorff, *Chem. Ing. Tech.* **47**, 125–128 (1975).
15. N. M. Prout and J. S. Moorhouse, eds., *Modern Chlor-Alkali Technology*, Vol. 4, Society of Chemical Industry, Elsevier Applied Science Publishers, Ltd., London, 1990, pp. 40–42, 97, and 178.
16. *Chlorine Engineers' Membrane Electrolyzer*, brochure no. EC-T002-3, Chlorine Engineers Corp., Ltd., Tokyo, 1991.
17. M. Yoshida and Y. Naoki, *The Electrochemical Society Meeting*, abstract #364, Los Angeles, May 1989.
18. A. Bjareklint, L. Carlsson, and B. Sandergrén, in Ref. 12, pp. 197–205.
19. R. E. Connick and Y-T. Chia, *J. Am. Chem. Soc.* **81**, 1280 (1959).
20. W. Latimer, *Oxidation Potentials*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1952.
21. D. V. Kokoulind and L. I. Krishtalik, *Sov. Electrochem.* **7**, 325 (1971).
22. B. V. Tilak, K. Viswanathan, and C. G. Rader, *J. Electrochem. Soc.* **128**, 1228–1232, (1981).
23. A. T. Kuhn, *Industrial Electrochemical Processes*, Elsevier Publishing Co., Amsterdam, The Netherlands, 1971, Chap. 3.
24. J. S. Sconce, ed., *Chlorine*, ACS Monograph, Reinhold Publishing Co., New York, 1962.
25. T. C. Jeffery, P. A. Danna, and H. S. Holden, eds., *Chlorine Bicentennial Symposium*, The Electrochemical Society Softbound Series, Princeton, N.J., 1974.
26. Ref. 4, pp. 2981–2983.
27. **U.S. Pat. 4,410,411** (Oct. 18, 1983), R. W. Fenn III and co-workers (to Oxytech Systems, Inc.).
28. M. P. Grotheer, *The Electrochemical Society Meeting*, Extended Abstracts, Minneapolis, Minn., May 1981, pp. 1010–1012.
29. T. F. Florkiewicz and L. C. Cudin, in T. C. Wellington, ed., *Modern Chlor-Alkali Technology*, Vol. 5, Society of Chemical Industry, Elsevier Applied Science Publishers, Ltd., London, 1992, pp. 209–221.
30. *Alkaline Chloride Electrolysis by the Diaphragm Process*, System Hooker, Uhde GmbH, Dortmund, Germany, 1985.
31. W. C. Gardiner, in T. R. Beck and co-workers, eds., *Electrochemical Contributions to Environmental Protection*, The Electrochemical Society Symposium Series, Princeton, N.J., 1972, p. 16.
32. *North American Chlor-Alkali Industry Plants and Production Data Book*, pamphlet 10, The Chlorine Institute, Washington, D.C., Jan. 1989.
33. **U.S. Pat. 3,770,567** (June 6, 1971), W. Gratt (to E.I. du Pont de Nemours & Co., Inc.).
34. D. Bergner, *Membrane Electrolyzers for the Chlor-Alkali Industry*, Dechema-Monographs, Vol. 123, VCH Verlagsgesellschaft, New York, 1991, pp. 113–131.
35. Ref. 15, papers 7, 11, 12, 13, 15, 16, 17, 18, and 20.
36. H. Shiroki, T. Hiyoshi, and T. Ohta, in Ref. 29, pp. 117–129.
37. R. F. Bradley, A. Leder, and T. Sasano, *Sodium Chlorate, Chemical Economics Handbook*, SRI International, Menlo Park, Calif., Oct. 1991.
38. M. P. Grotheer and E. H. Cook, *Electrochem. Technol.* **6**, 221 (1968).
39. N. Munichandraiah and S. Sathyanarayana, *J. Applied Electrochem.* **17**, 33–48 (1987).
40. **Jpn. Kokai 51 104,499** (Sept. 16, 1976), K. Nakamura (to Japan Metals and Chemicals Co., Ltd.).
41. **U.S. Pat. 4,405,419** (Sept. 20, 1983), M. Misawa and co-workers (to Japan Metals and Chemicals Co., Ltd.).
42. **U.S. Pat. 4,818,354** (Apr. 4, 1989), E. Preisler, J. Holzem, and G. Mietens (to Hoechst AG).
43. **U.S. Pat. 4,997,531** (Mar. 5, 1991), M. Y. Sagashi and R. S. Ichikawashi (to Japan Metals and Chemicals Co., Ltd.).
44. K. V. Kordesch, ed., *Batteries*, Vol. 1, Marcel Dekker, Inc., New York, 1974, pp. 469–475.
45. B. R. A. Wood and co-workers, *Progress in Acoustic Emissions V*, The Japanese Society for Non-Destructive Inspection, Tokyo, 1990, pp. 285–292.
46. C. B. Ward and D. A. J. Swinkels, *Chem. Aust.*, –499 (Nov. 1991).
47. E. Preisler, *Progress in Batteries and Battery Materials*, Vol. 10, International Battery Association, Brunswick, Ohio, 1991, pp. 1–22.
48. D. Glover, B. Schumm, Jr., and A. Kozawa, eds., *Handbook of Manganese Dioxides, Battery Grade*, International Battery Material Association, Brunswick, Ohio, 1989, p. 11.
49. B. Heydon, H. Schwendener, and M. Tashiro, *Hydrogen*, CEH Product Review, SRI International, Menlo Park, Calif., 1990.
50. Ref. 23, Chap. 4.
51. Ref. 11, pp. 33–50.
52. J. T. Anderson, *Chem. Proc. (London)* **12**, 513 (1966).
53. A. E. Zdansky, *Dechema Monog.* **33**, 92 (1959).
54. G. G. Scherer, T. Momose, and K. Tomiue, *J. Electrochem. Soc.* **135**, 3071–3073 (1988).
55. Y. Ogata, M. Yasuda, and F. Hine, *J. Electrochem. Soc.* **135**, 2976–2981 (1988).
56. L. F. Allright and co-workers, *Energy Prog.* **8**, 22–27 (1988).
57. J. F. McElroy, *J. Power Sources* **36**, 219–233 (1991).
58. Ref. 23, Chap. 5, pp. 161–165.
59. *Chem. Eng. News*, 4–6 (Apr. 3, 1989).
60. *Business Week*, 86–87 (Apr. 10, 1989).
61. *Newsweek*, 66 (May 1, 1989).

62. *Chem. Eng. News*, 23–24 (May 25, 1992).
63. Ref. 23, Chap. 1.
64. J. Dykstra and co-workers, *Ind. Eng. Chem.* **47**, 883 (1955).
65. S. Koyama, T. Maeda, and K. Okumura, *Z. Anorg. Allg. Chem.* **540–541**, 117–134 (1986).
66. N. Watanabe, T. Tojo, and Y. B. Chong, *Tanso* **145**, 250–255 (1990).
67. Ref. 23, p. 504.
68. Ref. 11, pp. 231–234.
69. **U.S. Pats.** **2,908,620** (Oct. 13, 1959), **3,062,734** (1962), and **4,085,191** (Apr. 18, 1978), (to Carus Chemical Co.).
70. G. R. Popov, *Khim. Ind. (Sofia)* **40**(4), 166–169 (1968).
71. T. Kaczmarek, J. Chwalczyk, and B. Kolomyjek, *Chemile* **25**(2) 58 (1972).
72. V. D. Maslov, Yu. S. Gurinov, and Yu. Ya. Kharitonov, *Otkrytiya. Izobret. Prom. Obrabztsy Tovarnye Znaki* **3**, 131 (1980).
73. **U.S. Pat.** **3,986,941** (Oct. 19, 1976), T. Okabe and co-workers (to Nippon Chemical Industry Co., Ltd.).
74. **U.S. Pat.** **5,062,930** (Nov. 5, 1991), J. R. Dillon and J. D. O'Connor (to Shipley Company Inc.).
75. Ref. 11, pp. 221–230.
76. P. W. Sherwood, *Chem. Ing. Tech.* **32**, 459 (1960).
77. B. Heydorn, A. Leder, and Y. Ishikawa, *Hydrogen Peroxide*, CEH Product Review, SRI International, Menlo Park, Calif., June 1992.
78. *Chem. Eng.*, 32–37 (June 20, 1988).
79. *Chem. Week*, 36–37 (Jan. 27, 1988).
80. *On-Site Peroxide Generation Systems*, brochure, HD Tech Inc., Etobicoke, Ontario, Canada, 1991.
81. Ref. 4, pp. 2984–3000.
82. Ref. 23, Chap. 6.
83. G. Milazzo, *Electrochemistry, Theoretical Principals and Practical Applications*, Elsevier Publishing Co., Amsterdam, The Netherlands, 1963, pp. 439–482.
84. R. H. Newnham, *J. Applied Electrochem.* **22**(2), 116–124 (1992).
85. **Can. Pat.** **1,190,184** (July 9, 1985), H. B. Beer and J. M. Hinden (to Eltech Systems Corp.).
86. **Ger. Pat.** **3,731,285** (Apr. 6, 1989), E. Wenk and K. Koziol (to C. Conradty Nuernberg GmbH & Co., K-G.).
87. G. Ferron and P. F. Duby, in F. Hine and co-workers, eds., *Performance of Electrodes for Industrial Electrochemical Processes*, The Electrochemical Society Symposium PV89-10, Los Angeles, Calif., 1989, pp. 259–277.
88. **U.S. Pat.** **5,047,133** (Sept. 10, 1991), R. J. Allen (to Metallgesellschaft AG).
89. **U.S. Pat.** **4,478,697** (Oct. 23, 1984), T. W. Coleman and R. A. Griffin (to Kerr-McGee Chemical Corp.).
90. S. E. Afifi and co-workers, *J. of Metals* **44**(1), 32–34 (1992).
91. K. J. Cathro, *Chem. Aust.* **58**(11), 490 (1991).
92. P. L. Claessens and G. B. Harris, eds., *Proceedings International Symposium on Electrometallurgical Plant Practices*, Pergamon Press, New York, 1990, pp. 115–127.
93. N. A. Gardner and F. C. Walsh, in Ref. 12, pp. 225–258.
94. **U.S. Pat.** **4,632,738** (Dec. 30, 1986), M. J. V. Beathe, W. G. Bacon, and R. Raudsepp (to Great Central Mines Ltd.).
95. **Eur. Pat. Appl.** **EP 170,632** (Feb. 5, 1986), E. Diaz Nogueira.
96. T. J. O'Keefe, *J. Electroanal. Chem. Interfacial Electrochem.* **168** (1–2), 131–146 (1984).
97. G. W. Barton and A. C. Scott, *J. Appl. Electrochem.* **22**(2), 104–115 (1992).
98. R. Keller, J. W. Burgman, and P. J. Sides, in L. G. Boxall, ed., *Light Metals 1988*, The Metallurgical Society, Inc., Warrendale, Pa., 1988, pp. 629–631.
99. R. L. Rolf and R. W. Peterson, in R. D. Zabreznik, ed., *Light Metals 1987*, The Metallurgical Society, Inc., Warrendale, Pa., 1987, pp. 209–213.
100. H. Bentzen, *Light Metals 1991*, The Metallurgical Society, Inc., Warrendale, Pa., 1991, pp. 741–747.
101. **Can. Pat.** **2,003,660** (July 9, 1990), D. W. Townsend.
102. H. L. King, *Energy Res. Abstr.* **15**(21), Abstr. No. 47130 (1990).
103. **Eur. Pats.** **60,579**, **50,680**, and **50,681** (May 5, 1982), K. Koziol and co-workers (to C. Conradty Nuernberg GmbH & Co., K-G.).
104. T. R. Beck and R. T. Ruggeri, in H. Gerischer and C. W. Tobias, eds., *Advances in Electrochemistry and Electrochemical Engineering*, Vol. 12, John Wiley & Sons, Inc., New York, 1981, pp. 263–354.
105. M. Sittig, *Sodium: Its Manufacture and Uses*, Reinhold Publishing Co., New York, 1956, pp. 29–34.
106. Z. Qiu and Q. Li, in Ref. 98, pp. 775–780.
107. **U.S. Pat.** **4,203,819** (May 20, 1980), S. A. Cope (to E.I. du Pont de Nemours & Co.).
108. **U.S. Pat.** **4,533,442** (Aug. 6, 1985), R. Neelameggham and J. C. Priscu (to AMAX, Inc.).
109. T. Takei, *J. Appl. Electrochem.* **9**(5), 587–593 (1979).
110. *A Survey of Electrochemical Metal Winning Processes*, ANL OEPM-79-3, Battelle Columbus Laboratory, Argonne, Ill., Mar. 1979.
111. O. G. Sivillotti, in Ref. 98, pp. 817–822.
112. T. Ishikawa, T. Narita, and S. Konda, *Proceedings of the Electrochemical Society 7th International Symposium on Molten Salts*, Vol. 90-17, 1990, pp. 600–610.
113. G. M. Rao, *J. Electroanal. Chem. Interfacial Electrochem.* **249**(1–2), 191–203 (1988).
114. G. J. Kipouros and D. R. Sadoway, *Adv. Molten Salt Chem.* **6**, 127–209 (1987).
115. Ref. 23, p. 208.
116. M. M. Wong, D. E. Couch, and D. A. O'Keefe, *J. Metals* **21**(1), 43–45 (1969).
117. W. Samhaber, in Ref. 12, pp. 207–223.
118. E. M. Paton and K. Scott, *The Electrochemical Society Meeting*, abstract #548, Montreal, Quebec, May 1990.
119. S. K. Mishra and D. McCarty, *The Electrochemical Society Meeting*, abstract #561, Montreal, Quebec, May 1990.
120. F. C. Walsh and D. E. Saunders, *J. Photograph. Soc.* **31**, 35 (1983).
121. S. Stucki and co-workers, *J. Applied Electrochem.* **21**, 99–104 (1991).
122. T. C. Franklin and co-workers, *J. Electrochem. Soc.* **138**, 2285–2288 (1991).
123. T. C. Franklin, J. Darlington, and W. K. Adenuji, *J. Electrochem. Soc.* **137**, 2124–2126, (1990).
124. **U.S. Pats.** **4,256,552** (Mar. 17, 1981), **4,324,635** (Apr. 13, 1982), **4,334,968** (June 15, 1982), **4,596,648** (June 24, 1986), and **4,804,449** (Feb. 14, 1989), C. T. Sweeney.
125. N. A. Gardner and F. C. Walsh, in Ref. 12, pp. 225–258.
126. R. J. Horvath, *PC FAB*, 38–40 (June 1992).
127. Brochure EIC-6350, Eltech International Corp., Fairport Harbor, Ohio, 1991.
128. **U.S. Pat.** **3,859,195** (Jan. 7, 1975), J. M. Williams (to E.I. du Pont de Nemours & Co., Inc.).
129. P. M. Robertson and co-workers, *Chem. Ind.*, 459 (July 1, 1978).
130. G. B. Adams, R. P. Hollandsworth, and D. N. Bennion, *J. Electrochem. Soc.* **122**, 1043–1048 (1975).
131. Brochure EIC-6356, Eltech International Corp., Fairport Harbor, Ohio, 1992.
132. C. Commimellis and C. Pulgarin, *J. Appl. Electrochem.* **21**, 703–708 (1991).
133. R. F. Schwab and W. H. Doyle, *Electrochem. Technol.*, 228–236 (May–June 1967).

134. S. G. Epstein and R. E. Miller, in Ref. 99, pp. 693–698.
135. R. E. Knowlton, in Ref. 15, pp. 45–50.
136. Ref. 15, papers 5, 6, and 19.
137. Ref. 29, papers 1, 4, 10, 19, and 20.
138. Ref. 99, pp. 623–688.
139. *Chem. Eng.* **99**(6), 26–27 (June 1992).
140. *Occupational Injuries and Illnesses in the United States by Industry, 1990*, bulletin 2399 Bureau of Labor Statistics, Washington, D.C., Apr. 1992.

General Reference

Light Metals, The Metallurgical Society, Warrendale, Pa., published annually.

Morris P. Grotheer
Kerr-McGee Corporation



ORGANIC

An electroorganic reaction is a heterogeneous process having similarities to heterogeneous catalysis (see CATALYSIS). In both cases selectivity is influenced by materials used, ie, the electrode, mass transport, mixing rates, and surface and bulk regimes. The development of a viable electroorganic reaction benefits from the application of chemical engineering methods. Electrochemical engineering, which began to take shape in the mid-1950s, is an interdisciplinary field (1) covering such factors as current distribution and interfacial electrochemistry as well as chemical engineering. There are a number of texts covering electrochemical engineering and technology (2–14).

The electrochemical cell simultaneously provides for oxidation and reduction processes, which occur at the anode and cathode, respectively. An organic molecule on the surface of an electrode can undergo electron removal (anode) or electron addition (cathode), resulting for uncharged organic reactants in the formation of a radical cation or radical anion, respectively. These reactive species, at essentially ambient temperatures, are open to a wide range of reaction paths depending on the species, medium, and adsorption at the electrode. Reviews of the wide range of electroorganic reactions investigated are available (15–24).

Oxidation and reduction reactions can be carried out using reformer hydrogen and oxygen from the air. To decide when electroorganic synthesis is likely to be a viable option for a desired product, some opportunity factors are use of cheaper feedstock; elimination of process step(s) or a difficult reaction; avoidance of waste disposal, toxic materials, and or ability to recycle reagent; and ability to obtain products from anode and cathode.

For a profitable electrochemical process some general factors for success might be listed as high product yield and selectivity; current efficiency >50%; electrolysis energy <8 kWh kg product; electrode, and membrane in divided cells, lifetime >1000 hours; simple recycle of electrolyte having >10% concentration of product; simple isolation of end product; and the product should be a key material and or the company should be comfortable with the electroorganic method.

Attempts to emulate the few large-scale electroorganic processes, eg, production of tetraalkyllead and adiponitrile, have often led to frustration. There are, however, numerous applications forthcoming in medium- to small-scale processing. Especially attractive on this scale is the pharmaceutical fine chemical or high value added chemical synthesis (see FINE CHEMICALS). In these processes multistep reactions are common, and an electroorganic reaction step can aid in process simplification. Off the shelf lab electrochemical cells, which have scaled-up versions, are also available. The materials of construction for these cells are compatible with most organic chemicals.

Cell Design

Materials.

Cell Structure. Because electroorganic syntheses in the main are carried out at close to room temperature, a wide range of plastics are available for cell construction. A full range of readily available polymers, eg, polyethylene, poly(vinylidene fluoride) (PVDF), polypropylene, polytetrafluoroethylene (PTFE), and other thermoplastic materials, is used for cell frames or vessels, and elastomers (qv), eg, neoprene, Viton, ethylene–propylene–diene monomer (EPDM), and PTFE encased, are used to seal cell parts. Because the cell contents are likely to have solvent action, materials used in cell construction should be tested for long-term stability in the electrolyte mix. Testing should consider not only the solubility of the materials, but also dimensional stability to avoid problems at tracks and joints. In addition, the materials should be screened for conductivity, because some polymers and elastomers have conductive fillers that preclude use for insulated cell parts. Whereas machining can be used for fabrication, injection molding is often the preferred method for multiple manufacture. Some form of external steel reinforcing can be used if strength is required.

Electrodes. At least three factors need to be considered in electrode selection as the technical development of an electroorganic reaction moves from the laboratory cell to the commercial system. First is the selection of the lowest cost form of the conductive material that both produces the desired electrode reactions and possesses structural integrity. Second is the preservation of the active life of the electrodes. The final factor is the conductivity of the electrode material within the context of cell design. An in-depth discussion of electrode materials for electroorganic synthesis as well as a detailed discussion of the influence of electrode materials on reaction path (electrocatalysis) are available (25,26). A general account of electrodes for industrial processes is also available (27).

Selection. The widely used cathode materials include Hg, Pb, Al, Zn, Ni, Fe, Cu, Sn, Cd, and C. Because of mechanical inconvenience, mercury is not an attractive electrode material for large cells. The preferred material is lead or an amalgam. Because Pb is soft and has a tendency to deform, however, it presents some mechanical problems. The problems can be overcome by hot dip or electroplating on steel, copper, or other rigid base material. Alternatively, to harden the lead and improve the dimensional stability, it may be possible to introduce a few percent of an alloying element such as silver, antimony, calcium, or copper. Some caution should be used with alloying elements because these can markedly change by-product formation. Given its low cost and high mechanical strength, steel is often used in electrolytic processes in which the cathode is the counter electrode and hydrogen evolution is desired. If a lower hydrogen overvoltage is needed, nickel or the more expensive Raney Ni coated cathode, as developed for chlor-alkali cells by Dow and Eltech, may be used in place of steel.

Corrosion problems are common for anode materials, thus the list is shorter than for cathodes. If the cathode product is of sufficient value, stoichiometric corrosion may be accepted. Often used anodes include Pt, C, and PbO₂ (for fabrication, see Ref. 28). Ag, Au, Cu, Fe, and Ni anodes might find application for organics more easily oxidized. Fe and Ni are relatively stable and have been widely used in high pH aqueous electrolytes. Platinum-clad or platinum-plated titanium or hard carbon (29) can be used for an anodic process such as the Kolbñ synthesis, which gives the best results on smooth platinum. Whereas it possesses some fabrication problems, magnetite is an oxide that could be used technically as an oxygen evolver. Methods of preparing this anode have been reviewed (30). The plasma spray technique is attractive (see PLASMA TECHNOLOGY). Because of magnetite's ceramic nature, fabricating the anodes as a coating on titanium is probably best.

Generally, steel or nickel is used in making anodes for oxygen evolution in alkaline electrolytes. Graphite or nickel is frequently employed as anodes for anodic halogenation of organics. The approach to anode construction has taken a shift following the invention of the so-called dimensionally stable anode (DSA) for chlorine evolution (25) (see ALKALI AND CHLORINE PRODUCTS). The Beer process uses a coating of ruthenium oxide [12036-10-1], RuO₂, and titanium dioxide [13463-67-7], TiO₂, on titanium to produce the electrode for chlorine. However, this electrode is not so long-lived as an oxygen generator. A better formulation is iridium dioxide [12030-49-8], IrO₂, and TiO₂ on titanium (or tantalum). Even though it works well as an oxygen evolver in inorganic systems, the IrO₂-based anode has shown a short life in the presence of organics, even at a low concentration. Companies that offer coated anodes and may include noble metal on titanium are Eltech, Engelhard, and Electrocatalytic, Inc. in the United States; ICI in the United Kingdom; Conradty in West Germany; and Permelec in Europe (see METAL ANODES).

Stability. In general, the best view of the electrode is as a catalyst in the process. Thus it should be both costed and assessed accordingly, eg, fractions of an electrode per kg of product. With time, the nature of the active electrode surface is likely to change either chemically or physically. The leaching out of a component from an alloy can change the electrode composition as can surface contamination from an insoluble organic by-product or electrodeposition of extraneous metals. In laboratory experiments using short runs and reagent-grade chemicals, electrode surface contamination is not always recognizable as a problem. However, it may prove to be a primary consideration in the development of a commercial process. Electrode surface contamination can be

circumvented by: (1) periodic mechanical and or chemical cleaning; (2) current reversal; (3) stringent raw material specifications with regard to problem impurities; (4) purging and processing a recycle electrolyte stream; or (5) replacing electrodes.

Either corrosion or erosion may result in mechanical deterioration of the electrode surface. To minimize such problems, electrode materials should be chosen with care. Pourbaux diagrams ie, plots of potential vs pH, point to the conditions that might lead to corrosion (25). Because the diagrams for metal–water alone can only give an approximate idea of corrosion conditions, the Pourbaux diagram should be generated to include all components of the electrolyte (31). The open circuit should be included in the possible corrosion conditions. When power is removed from the cell, such as for maintenance, it must be known how long electrolytes can remain inside the cells without causing damage. The intermediates of the organic electrode process may also cause corrosion. The respective radical ions may attack both anode and cathode to form organometallics. High rates of circulation of electrolyte containing solid particles of reaction products or electrode corrosion materials may cause erosion. This problem can be contained by removal of the solids external to the cell. Possible dislocation of electrode coatings can occur, and new bonding methods may have to be explored.

Conductivity. Electrical conductivity is important and should be arranged to be as high as possible. Energy losses and the consequent heating of electrode and electrical connectors are to be avoided. For cell design using a monopolar electrode configuration, pure electrode materials require a resistivity of less than 10^{−4} Ωcm (26). This helps not only to avoid energy losses, but also to maintain a uniform current distribution (26,32). Because it involves only short current paths through the electrode materials, bipolar design is better for more resistive electrodes such as oxides and conductive polymers.

Diaphragms. In many electroorganic systems, the use of a diaphragm to separate anolyte and catholyte is necessary to minimize the side reactions caused by the incompatibility of a cell's two half-cell reactions. In the development of a commercially feasible process, the selection of a diaphragm for large-scale cells often presents a serious obstacle. The preferred electrolysis diaphragm would have low cost, low electrical resistance, limited interdiffusion of anolyte and catholyte constituents other than carrying ions, long operating life, good dimensional stability, and be insusceptible to plugging and fouling. The probability of finding all of these qualities in a single material is, unfortunately, small. Diaphragm materials are available in two types: permeable and permselective. The permeable material is a porous matrix offering a limited barrier to all species in an electrolyte, and having little or no change in electrolyte transport numbers. Because of significantly lower cost, the permeable materials may be preferable to the permselective membranes. Permselective materials, or ion-exchange membranes, allow the passage of ions of one charge with a high degree of exclusion of ions of the opposite charge. Cation-exchange membranes allow greater mobility to cations than anions. Therefore, the bulk of the current transfer is the movement of cations from the anolyte to the catholyte. If movement of cations is to be suppressed, anion-exchange membranes are similarly used (see ION EXCHANGE; MEMBRANE TECHNOLOGY).

Permeable Diaphragms. In those cases when a small amount of flow between the anolyte and catholyte can be tolerated or perhaps is desired, a permeable separator is entirely satisfactory. Alundum cups and glass frits have been used in simple laboratory cells. In production cells, however, the fragile alundum and other forms of unglazed porcelain are not very practical (see ENAMELS, PORCELAIN OR VITREOUS). A variety of porous media, available in a range of porosities, serve as alternatives. These have been manufactured for separators or filters in electrical batteries. Available materials that might be used are porous thermoplastics of polyethylene or polypropylene, PVC, PTFE, polycarbonate, or isoprene rubber; filter cloth of glass, polyethylene or polypropylene, PVC, and PTFE; and nonwoven fabrics (qv) of polyethylene and polypropylene, and copolymer of vinyl chloride and acrylonitrile. Related to the permeability, porosity, pore size, and thicknesses are electrical and diffusional resistance of porous materials. Methods of evaluating and correlating the electrical conductivity of different porous media have been described (33). The hydrophobic character of some membranes can make them unsuitable for many applications, and too high a current density on a membrane can lead to local boiling of the electrolyte.

Permselective Diaphragms. There has been considerable activity in development of permselective diaphragms or ion-exchange membranes, especially for electrodialysis and the chlor-alkali processes. Several general reviews have been published on commercially available membranes (27,34,35). One method of preparing thin ion-exchange membranes involves the casting of a finely divided powder of ion-exchange resin in a plastic binder. Because of their heterogeneous character, these membranes have several disadvantages. Relatively large channels usually exist through the matrix between resin particles leading to high rates of electroosmotic water solvent transfer and molecular diffusion. On the other hand, homogeneous ion-exchange membranes are produced that are essentially a continuous polymeric structure. Several membrane manufacturers use a matrix of copolymer of styrene and divinylbenzene to produce a cross-linked structure. This is then aminated for anion exchange or sulfonated for cation exchange. For production of another series of membranes, the styrene and divinylbenzene can also be grafted to a polymer film and then aminated or sulfonated. To improve its mechanical properties, the polymer is frequently reinforced with a screen of Teflon, Saran, glass, or PVC.

Perfluorinated matrixes have been converted to ion-exchange membranes for the more rigorous environments of the chlor-alkali industry. The first of these, a sulfonated membrane for cation exchange given the trade name of Nafion ion-exchange membrane is marketed by Du Pont. These membranes have found application in electroorganic synthesis because of excellent chemical and physical stability as well as success in the chlor-alkali industry (36). A range of perfluorosulfonic membranes for various cell applications is available (37). Companies that provide commercial ion-exchange membranes are listed in Table 1. Asahi Glass, which has Flemion ion-exchange membrane, and Asahi Chemical of Japan produce perfluorocarboxylic acid-type cation-exchange membranes. Tokuyama Soda of Japan produces a fluorocarbon-based cation-exchange resin. Many of these perfluoro membranes are surface treated to perform well in high alkaline conditions.

Table 1. Sources of Commercial Ion-Exchange Membranes

Company	Type of membrane	Trade name ^a
Asahi Chemical, Tokyo	cation and anion	Aciplex
Asahi Glass Co., Tokyo	cation and anion	Selemon and Flemion ^F
E.I. du Pont de Nemours & Co., Inc., Wilmington, Del.	cation	Nafion ^F
Ionics, Watertown, Mass.	cation and anion	Ionics
RAI Research Corp., Long Island, N.Y.	cation and anion	Raipore
Sybron Chemicals Inc., Birmingham, N.J.	cation and anion	Ionac
Tokuyama Soda, Tokyo	cation and anion	Neosepta and Neosepta-F ^F
Tosoh Corp., Tokyo	anion	Tosflex ^F

^a The letter ^F indicates perfluorinated membranes.

RAI Research Corp. (U.S.) produces a range of anion- and cation-exchange membranes on polyethylene or polyfluoroethylene matrixes. Grafting techniques are used by RAI to build on the ion-exchange properties. Properties of ion-exchange membranes can be varied by altering polymer porosity, polymer composition, backing material, thickness, and exchange capacity. It is conceivable that a membrane can be tailor-made to fit the specific requirements of electroorganic synthesis. Examples of membrane evaluations are available (38,39). The references show that mechanical properties are important, closely followed by problems of unwanted transport of solvent or reactant. The latter not only constitutes a loss of materials, but also could result in the degradation of the counter electrode. Properties in commercial ion-exchange membranes range from 0.1–0.6 mm in thickness; 40–150 cm in width; 100–300 cm in length; 1.5–2.0 Ωcm resistance when measured in 0.5 N NaCl; 0.8–0.99 transport number when measured in 0.5 N NaCl; 10–60% water content; 2–20 kg cm² mullen burst strength; and \$50–500 m² cost.

Electrolyte. The ideal electrolyte, ie, the fluid part of the cell, for organic synthesis would give high solubility to the organic, possess good conductivity, have low cost, contain easy recovery and purification, and be noncorrosive. Quaternary ammonium salts provide many of the above criteria in aqueous systems. A concise compilation of solvents and salts used in electroorganic chemistry is available (40).

Transport Phenomena. Electrochemical reactions are heterogeneous and are governed by various transport phenomena, which are important features in the design of a commercial electroorganic cell system. As for other heterogeneous reactions, the electrochemical reaction is impacted by heat and

mass transport. The electrochemical reaction, however, is unique in that it also has charge-transport characteristics. More comprehensive works on transport phenomena are available (2–14,41).

Mass Transport. Probably the most investigated physical phenomenon in an electrode process is mass transfer in the form of a limiting current. A limiting current density is that which is controlled by reactant supply to the electrode surface and not the applied electrode potential (42). For a simple analysis using the limiting current characteristics of various correlations for flow conditions in a parallel plate cell, see Reference 43.

Scale-up of the limiting current in the direction of fluid flow for free convection at a vertical planar electrode is proportional to one over the $^{1/4}$ power of the distance up the electrode. The limiting current density becomes independent of electrode length for the horizontal configuration. An electrode under free convection, but having gas evolution, produces results that are similar to those seen under forced convection. In most electroorganic synthesis applications, forced convection environments provide the best cell conditions. Flow can be under either laminar conditions, ie, Reynolds No. (Re) <1000, or turbulent flow (Re >2500). When one considers the parallel plate electrode cell (a duct) for laminar flow conditions, the limiting current is inversely proportional to the $^{1/3}$ power of the distance along the electrode. Entrance effects are not accommodated, and a further correlation is required for this region.

Eddy diffusion as a transport mechanism dominates turbulent flow at a planar electrode in a duct. Close to the electrode, however, transport is by diffusion across a laminar sublayer. Because this sublayer is much thinner than the layer under laminar flow, higher mass-transfer rates under turbulent conditions result. Assuming an essentially constant reactant concentration, the limiting current under turbulent flow is expected to be independent of distance in the direction of electrolyte flow.

Methods proposed for improving mass-transfer rates in large-scale cells are (1) rotation of cylindrical or disk electrode including wiping of surface; (2) use of turbine or propeller agitators; (3) fluidized bed of electrode particles; (4) fluidized bed of nonconducting particles; (5) vibration of electrode; (6) gas sparging; and (7) external pumping of electrolyte in open channels or channels having turbulence promoters. In laboratory work, rotating-type electrodes are attractive. However, these run the risk of being an expensive solution to the mass-transfer problem in commercial cells. Costs emerge in the form of running, investment, maintenance, and crevice corrosion.

For many electrode reactions, it is just as important to transfer the products away from the vicinity of the electrode as it is to move the reactants to the electrode. As an example, cathodic reductions in aqueous solution generate base. This local shift in pH can lead to unwanted side reactions such as in the electrohydrodimerization of acrylonitrile (44). Here the formation of biscyanoethyl ether, the formation of which is pH controlled, was minimized by increasing the mass-transfer rate within the cell. In this example, increased mass transfer was achieved by increased electrolyte velocity. Using the same electrode reaction, Asahi employed turbulence promoters in the catholyte compartment (45), similar to those used in dialysis cells. As an approach to the mass-transfer problem in cell design for electroorganic synthesis, rotating cylinder electrodes have attracted some attention (46). A general discussion of mass-transfer problems for a range of cell designs is available (47).

Charge Transport. Side reactions can occur if the current distribution (electrode potential) along an electrode is not uniform. The side reactions can take the form of unwanted by-product formation or localized corrosion of the electrode. The problem of current distribution is addressed by the analysis of charge transport in cell design. The path of current flow in a cell is dependent on cell geometry, activation overpotential, concentration overpotential, and conductivity of the electrolyte and electrodes. Three types of current distribution can be described (48) when these factors are analyzed, a nontrivial exercise even for simple geometries (11).

Primary Current Distribution. As shown in Figure 1, current flow is normal to lines of equipotential. Primary current distribution is influenced only by geometric factors, and in most cells of this arrangement, results in high current density at an edge of an electrode. The primary current distribution at the edges of the electrode is theoretically infinite. Depending on the angle the insulator makes with the electrode, the current density at the edge is lower for an electrode not mounted flush in a cell (11), and for a crevice approaches zero. The tendency for corrosion to occur where plastic spacers meet an electrode is explained by this local depolarization.

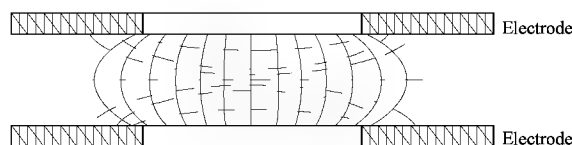


Fig. 1. Current flow () and electrical potential distribution () between two planar electrodes separated by an insulated channel.

Courtesy of Marcel Dekker, Inc., (49).

Secondary Current Distribution. When activation overvoltage alone is superimposed on the primary current distribution, the effect of secondary current distribution occurs. High overpotentials would be required for the primary current distribution to be achieved at the edge of the electrode. Because the electrode is essentially unipotential, this requires a redistribution of electrolyte potential. This, in turn, redistributes the current. Therefore, the result of the influence of the activation overvoltage is that the primary current distribution tends to be evened out. The activation overpotential is exponential with current density. Thus the overall cell voltages are not ohmic, especially at low currents.

Tertiary Current Distribution. The current distribution is again impacted when the overpotential influence is that of concentration. As the limiting current density takes effect, this impact occurs. The result is that the higher current density is distorted toward the entrance of the cell. Because of the nonuniform electrolyte resistance, secondary and tertiary current distribution are further complicated when there is gas evolution along the cell track. Examples of investigations in this area are available (50–52).

Heat Transfer. Cells are designed to minimize electrical power consumption. Heat removal, however, is probably a requirement because of the electrical resistive heating of the electrolyte. Although not economical, this becomes a convenient method of cell temperature control. Heat removal for the commercial electrolysis system is generally carried out by internal or external evaporative cooling; circulation of electrolyte through external heat exchangers; or internal cooling with coils, jackets, or tubes that may act as electrodes. Each of these methods has found some application in the commercial context. The standard chemical engineering practice (53,54) is followed for other aspects of cooling system design.

Reaction Engineering. Electrochemical reaction engineering considers the performance of the overall cell design in carrying out a reaction. The joining of electrode kinetics with the physical environment of the reaction provides a description of the reaction system. Both the electrode configuration and the reactant flow patterns are taken into account. More in-depth treatments of this topic are available (8,9,10,12).

Electrochemical Cells as Reactors. The electrochemical reactor readily parallels the chemical reactor of which there are two types used as ideal models. The first is the plug flow reactor (PFR) where reactants and products move through the reactor in a plug-like manner. No mixing occurs in the PFR, which presumes that concentration changes occur only in the direction of electrolyte flow. The second model is referred to as a back-mix or stirred tank reactor (STR). The STR assumes perfect mixing and uniform composition in all zones of the reactor. An E is inserted for the electrochemical version of these abbreviations to produce PFER and STER. Idealized models are not attained in practice: mixing occurs in the PFER, whereas imperfect mixing occurs in the STER. Thus reactors are often described as having either PFER or STER features.

The PFER and STER reactor types are influenced by mode of operation. The PFER can be used as shown (Fig. 2). Alternatively, a circulation loop can be added so that the inlet composition contains product. It is possible to shorten the length of the reactor and have higher mass-transfer rates using this latter configuration. The change in composition in these reactors is dependent on electrode length or flow rate. The plug flow model can be of use in estimation of electrode lengths Reference 56 provides a recent example of this reactor model analysis.

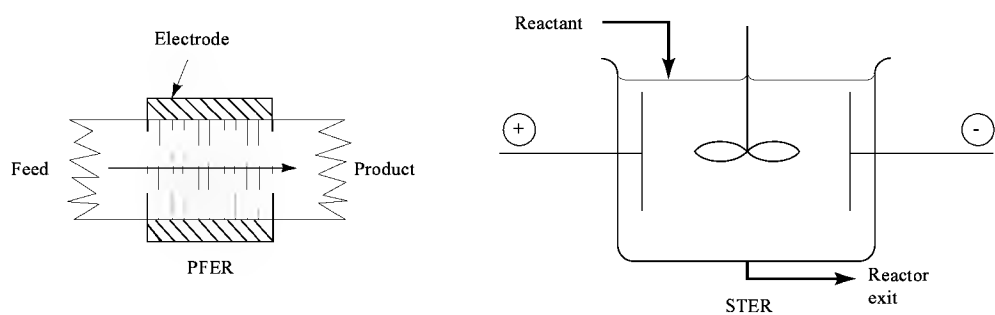


Fig. 2. Schematic models of a plug flow electrochemical reactor (PFER) and a stirred tank electrochemical reactor (STER).

Courtesy of Marcel Dekker, Inc. (55).

As shown in Figure 3, the STER can be configured for four modes of operation. A narrow gap divided cell is depicted in these schematic drawings. This narrow gap divided cell is shown in a circulation loop containing a surge tank and heat exchanger. Because of the dominant need for low cell resistance and good mass transfer through forced convection, this is probably the most likely general cell arrangement for the STER. Acceptable current distribution is provided by parallel plates. Figure 4 shows the reactor compositions for the various modes of operation. A single charge of reactant is made for the batch reactor. After sufficient reaction time, the charge, which is now product, is removed. As seen in Figure 4b, conversion depends on time. With the exception of the reactant composition in the batch being kept constant with time, the semibatch reactor configuration has the same feature. For the continuous STER (Fig. 4c), the compositions are constant with time. The first step in product refining with this arrangement is recovery of reactant for recycle back to the reactor. When continuous reactors are linked in a cascade, the final scheme for PFER is produced. This arrangement gives an overall performance similar to that of the PFER, such that each reactor is at a point on the curves of the PFER. Some work on modeling the STER is available (58).

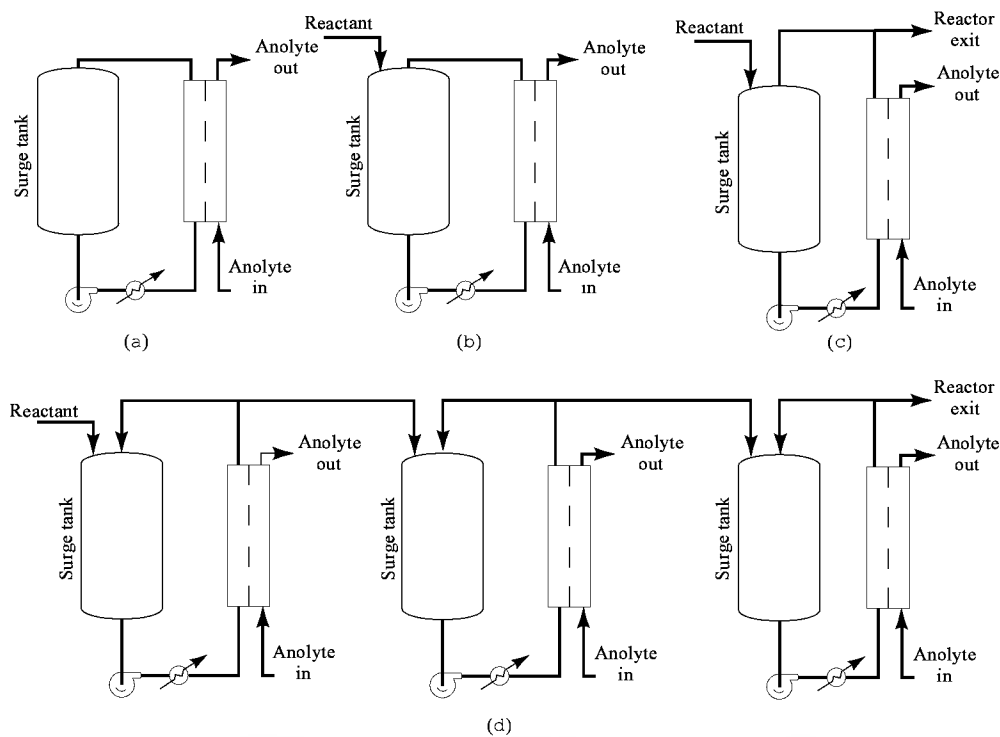


Fig. 3. Configurations for the stirred tank electrochemical reactor (STER). The surge tank contains both reactant and product. (a), Batch; (b), semibatch; (c), continuous; and (d), cascade.

Courtesy of Marcel Dekker, Inc. (55).

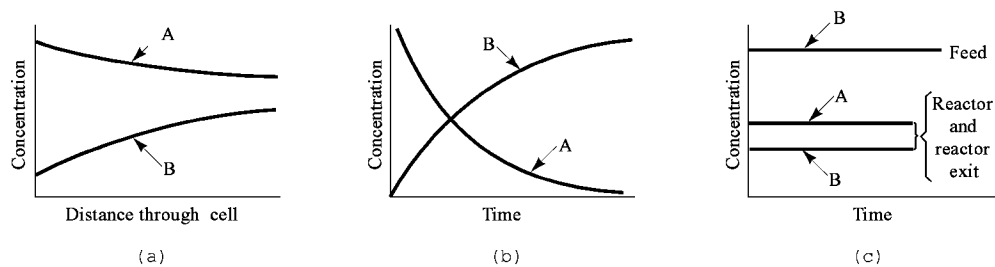


Fig. 4. Representations of reactor compositions for modes of cell operation where A represents product and B, reactant. (a), PFER; (b), batch STER; and (c), continuous STER.

Courtesy of Marcel Dekker, Inc. (57).

Economics and key characteristics of the electrode process govern the choice of reactor configuration and mode of operation. Where reaction rates have to be matched, engineering design can be critical in reaction schemes that involve electrochemical regeneration of a reactant. The regeneration of Mn^{3+} in the reaction sequence of butadiene to sorbic acid (59) is an example. Continuous operation is preferred for commercial cells. However, pharmaceutical products are usually made in batches so that materials can be segregated. Reactor choice is determined when a feature of the reaction significantly determines yield. The electrohydrodimerization of acrylonitrile to adiponitrile is an example. In this case, there is an optimum concentration of

acrylonitrile in the electrolyte for minimum by-product formation (60). Thus the continuous STER-type reactor is used. The product refining first goes through a reactant recovery step (61). Electrochemical cell design in the context of reaction engineering and electroorganic synthesis has been reviewed (62).

Scale-Up of Electrochemical Reactors. The intermediate scale of the pilot plant is frequently used in the scale-up of an electrochemical reactor or process to full scale. Dimensional analysis (qv) has been used in chemical engineering scale-up to simplify and generalize a multivariant system, and may be applied to electrochemical systems, but has shown limitations. It is best used in conjunction with mathematical models. Scale-up often involves seeking a few critical parameters. For electrochemical cells, these parameters are generally current distribution and cell resistance. The characteristics of electrolytic process scale-up have been described (63–65).

Scale-up is sometimes achieved by small multiples of reactors in standard chemical engineering. This is normal, and the multiples large, in the case of electrochemical cells. Because it gives reliable performance, this is an attractive form of scale-up; however, it loses the cost advantage of the large reactor. Options in hydraulic and electrical connection accompany multiple cells and are shown in Figures 5 and 6, respectively. The electrolyte may be run in parallel from one manifold to another in a given stack of cells. Alternatively, electrolyte can flow in series. Parallel flow has the general advantages of small temperature rise, lower gas fraction problems, low pressure drop, and low conversion per pass (some reactions are optimum at a constant reactant concentration). Parallel flow has the possible disadvantages, however, of nonuniform flow distribution, large volumetric flows, low conversion per pass (keeping recycle of reactant to a minimum), and higher bypass currents. Electrical connection to the cell bank can be monopolar or bipolar. Bipolar cell connection has the advantages of fewer busbar connections, more uniform current in each cell, better matching to rectifier equipment, lower voltage losses because of fewer connections, and no problem of current distribution down a plate because of plate resistance. The bipolar arrangement has the disadvantages of hazard present with high d-c voltages, disruption of a whole bank of cells in a cell malfunction, and current losses from bypass currents in the manifolds.

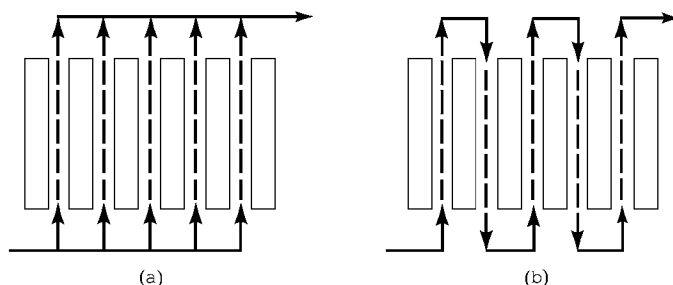


Fig. 5. Options for electrolyte flow through parallel plate cells: (a), parallel, and (b), series flow.

Courtesy of Marcel Dekker, Inc. (66).

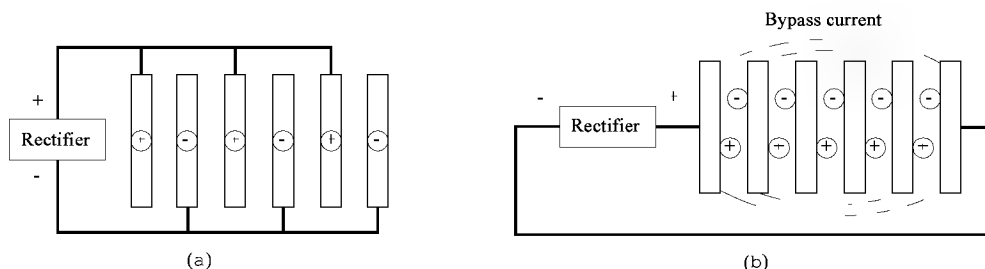


Fig. 6. Options for electrical connections to parallel plate cells: (a), monopolar and (b), bipolar connection.

Courtesy of Marcel Dekker, Inc. (67).

Over the years, a number of workers have modeled bypass currents which are also known as shunt, leakage, or parasitic currents, in bipolar cell stacks. These models have been based on analysis of an analogue resistance network of the cell bank. Although a simpler but more approximate model may be used (68), Reference 69 gives a comprehensive modeling. The fraction of leakage increases with approximately the square of the number of cells, but decreases with increased current density, because of the nonlinearity of overpotential and electrode length. Manifold design is a compromise between constricted high resistance channeling and tolerable pressure drop to achieve low current bypass at the cell inlet and exit. Corrosion of exposed metal may occur in this manifold region because of considerable voltage gradients.

Cell Geometries. Uniform electrode potential, short interelectrode gaps, and good mixing and mass transport benefit many electrochemical reactions. It is difficult to depart from the narrow spaced rectangular plates with turbulent flow electrolyte to achieve this. Reviews of electrolytic cell design are available in the literature (70,71). Several types of cell designs are reviewed herein. Design features in standard chemical engineering hardware are often used in various cell designs found in the literature. This is especially true if that equipment has a desired feature such as mass transfer, heat transfer, or gas absorption. The familiar filter press cells are made more sophisticated with internal manifolds like the plate heat exchanger, the fluidized and packed bed from catalytic reactors, etc.

Tank Cells. A direct extension of laboratory beaker cells is represented in the use of plate electrodes immersed into a lined, rectangular tank, which may be fitted with a cover for gas collection or vapor control. The tank cell, which is usually undivided, is used in batch or semibatch operations. The tank cell has the attraction of being both simple to design and usually inexpensive. However, it is not the most suitable for large-scale operation or where forced convection is needed. Rotating cylinders or rotating disks have been used to overcome mass-transfer problems in tank cells. An example for electroorganic synthesis is available (46).

Two-Dimensional Electrode Flow Cells. The simplest and least expensive cell design is the undivided parallel plate cell with electrolyte flow by some form of manifold. Electrical power is monopolar to the cell pack (72). An exploded view of the Foreman and Veatch cell is shown in Figure 7. Note that electrolyte flow is in series and that it is not easily adapted for divided cell operation.

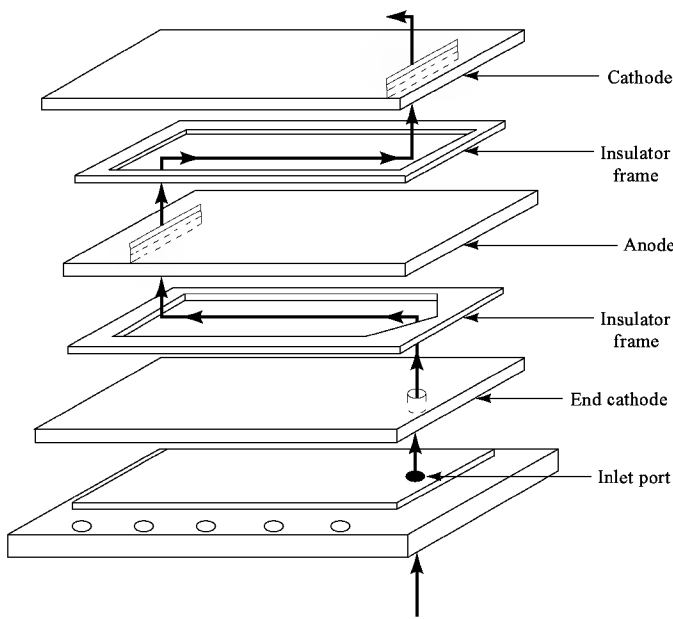


Fig. 7. Foreman and Veatch cell (73).

A possible problem of sealing the electrolyte path is found in the Foreman and Veatch cell. This can be avoided by placing the cells in a vessel. The best known example of this is the Beck and Guthke cell shown in Figure 8 (74). The cell consists of a stack of circular bipolar electrodes in which the electrolyte is fed to the center and flows radially out. Synthesis experience using this cell at BASF has been described (76). This cell exhibits problems of current by-pass at the inner and outer edge of the disk cells. Where this has become a serious problem, insulator edges have been fitted. The cell stack has parallel electrolyte flow; however, it is not readily adaptable to divided cell operation.

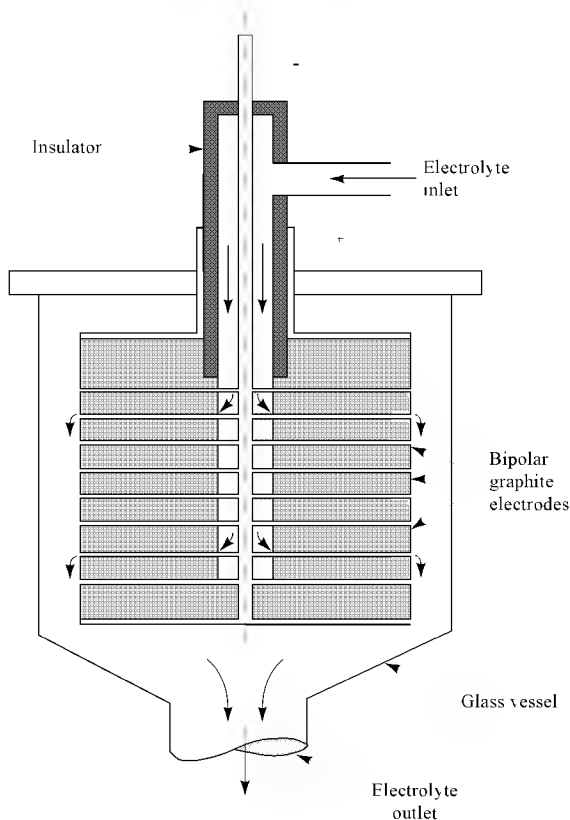


Fig. 8. BASF capillary gap cell (75).

Cell design work at Monsanto has also moved in the direction of vessel enclosed electrode stacks. As shown in Figure 9, one design (77) utilizes square rather than circular electrodes set in a cylindrical vessel. Manifolds are formed by the four segment cavities. To minimize current bypass in the bipolar operation, plastic electrode spacers insulate the cell edges. The advantages of this cell over the disk stack cell are uniform electrolyte flow and square electrodes that can be economically cut from sheets. Adaptation for divided cell operation (79) provides an interesting variation of this cell design. In this mode, the four manifolds are paired to give anolyte and catholyte flow paths. Alternate electrodes are replaced by diaphragm material. A further variation of the stacked electrode system is achieved by placing the electrodes vertically in the cylindrical vessel (80).

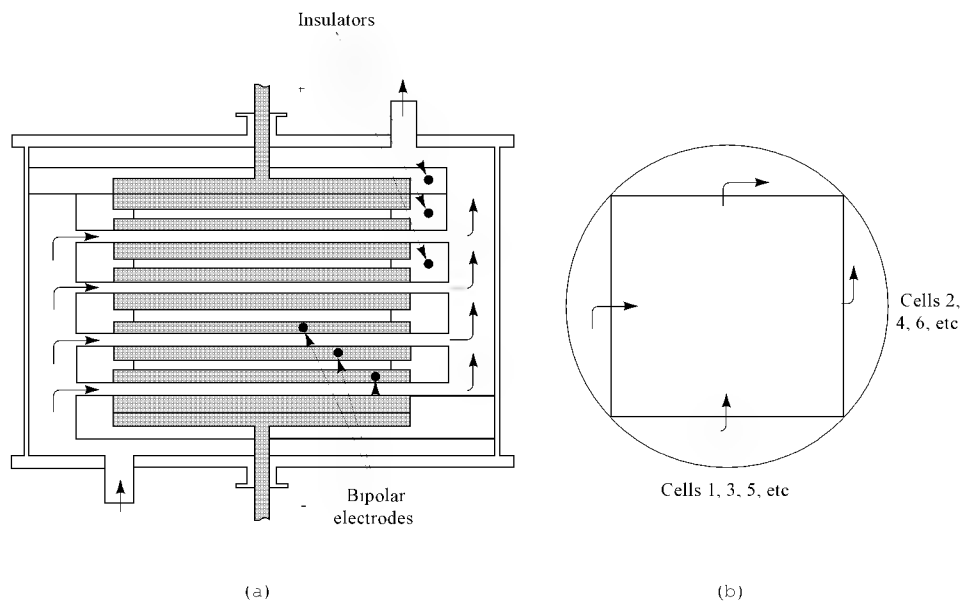


Fig. 9. Monsanto stack cell. (a), Side view of cell stack; (b), top view of flow paths across cells (78).

The divided cells often follow the plate and frame construction. An example of a cell used at Monsanto can be found in Reference 81. Plate and frame cells have also been used by other workers, including BASF (76). In addition, there are commercially available cells.

Three-Dimensional Electrode Flow Cells. Three-dimensional electrode flow cells have attracted attention for a number of years as electrode systems that give a high electrode area to cell volume ratio. In addition, these can have good mass-transfer characteristics. These cells parallel heterogeneous catalytic reactors in terms of being packed or fluidized beds of electrode material. Each may also function as monopolar or bipolar electrodes. The current distribution on the electrode surface is nonuniform for both of these electrical configurations. This may be undesirable. Electrolyte pressure drops through an electrode bed can be large. The possibility of handling electrode material as a process stream is an attractive feature of the packed-bed particulate electrode. The Nalco Chemical process for the manufacture of tetraalkyllead has done this on a large scale (82).

Commercially Available Cells. A significant obstacle to electrode process development was the design and construction of the electrochemical cell. This has changed because a number of companies fabricate a range of cells for general use in electroorganic synthesis. These are parallel plate cells that may be divided or undivided. All use electrolyte recirculation for convection, and may include turbulence promoters to enhance mass transport. They follow the plate and frame design with external or internal electrolyte manifolding. A variety of electrode materials is offered, eg, pure metals, PbO₂, DSA, graphite, and alloys, and the latest polymers are used for materials of construction, eg, polypropylene, polyvinylidenedifluoride, Teflon, and a range of elastomers. The peripherals of pumps, piping, etc, are also made available. Brochures on these cells are available from the respective companies, which also offer some consulting and support services.

ElectroCell AB. Cells developed through the cooperation of the Swedish National Development Co. (SU) and the foundation for Industrial Organic Chemistry at the University of Lund, Sweden, are marketed by ElectroCell AB (Akersberga, Sweden), jointly owned by SU and Investment AB Eken. The North American agent for these cells is the Electrosynthesis Co., Inc. (Lancaster, New York). A range of cell size is offered from a small benchtop unit to commercial-size modules. The statistics of these cells are outlined in Table 2, and an example is shown in Figure 10. The cells are of the plate and frame design using internal manifolding similar to a plate heat exchanger. With the exception of the Micro Flow cell, multiple cells are packed to form a module in which the electrolyte flows in parallel. The electrodes are electrically connected monopolar which avoid current bypass. The modules can be hydraulically fitted together as shown in Figure 10 for the ElectroSynCell model. Turbulence promoters are used to enhance mass transfer, and a cell divider is optional. The ElectroSynCell has been demonstrated in the production of glyoxylic acid from oxalic acid with current efficiencies of 84%. This, coupled with the low cell voltage, gives a power usage of 4.3 kWh/kg glyoxylic acid formed. For an indirect electrochemical process, the oxidation of the cerous to ceric ion has also been performed with a 90% current efficiency (83,84). The purchase price of these cells falls in the range of \$7,000–15,000 per square meter. These prices are dependent on quantity of cells and materials of construction.

Table 2. Characteristic ElectroCell AB Cells

Cell	Micro Flow	ElectroMp	ElectroSyn	ElectroProd
electrode area ^a , m ²	0.001	0.01	0.04	0.4
module area ^b , m ²	0.001	0.2	1.04	16
current density ^b , mA/cm ²	400	400	400	400
interelectrode gap, mm	3–6	6–12	5	0.5–4
electrolyte flow per cell, L/min	0.18–1.5	1–5	5–15	10–30
electrolyte velocity per cell, m/s	0.05–0.4	0.03–0.3	0.2–0.6	0.15–0.45

^a Value given is minimum.

^b Value given is maximum.

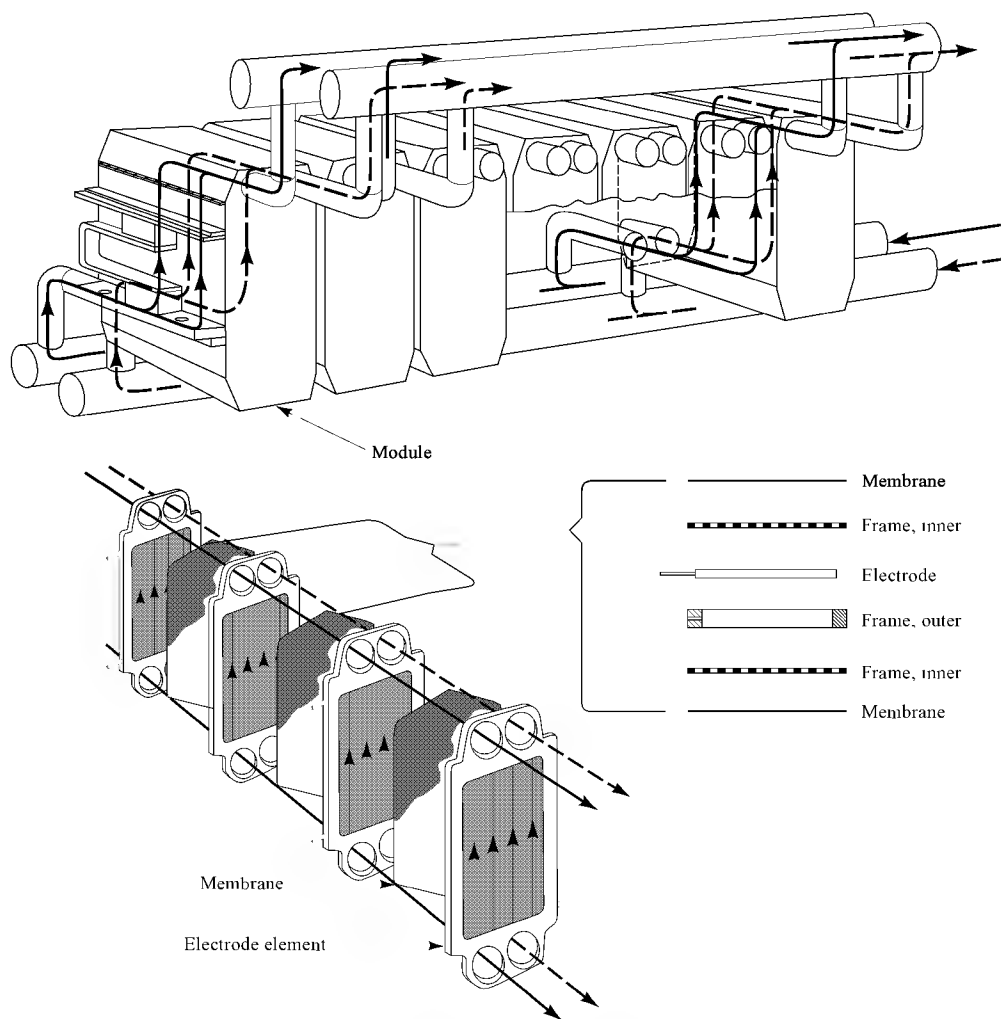


Fig. 10. The ElectroSynCell of ElectroCell AB.

ICI. In 1981 ICI made available the FM 21 cell for the chlor-alkali industry. FM 21 refers to filterpress, monopolar, 21 dm² membrane area. This successful cell was soon modified to the FM 21-SP (superior performance) for general electrolysis processes and particularly electroorganic synthesis by ICI C&P Ltd. in the UK. An overall view of the cell is shown in Figure 11. The cell can be either divided or undivided, and uses internal manifolding of electrolyte which is fed to the cell pack in parallel. Electrodes can be flat plate, bladed, or other enhanced surface area forms. The electrodes are spaced as low as 2 mm, and have a nominal projected area of 0.21 m². Electrical connection is monopolar by a copper bus attached along the long edge of the electrode. The cathode connection is on top and the anode on the underside. Several organic processes have been described as run in this cell (85,86). ICI has developed this design and other cells for lab to commercial-scale use. Some features of the range of cells available are summarized in Table 3. The benchtop cell version of the FM 21-SP is the FM01-LC unit. The FB series cells also have internal manifolding and are of a filter press design. However, they are bipolar in electrical connection, allowing a wider range of electrode materials to be used. Therefore, as with the SU cell, lab bench to pilot-plant-scale operation can be carried out in the same cell design. The cost of these cells is in the range of \$7,000–15,000 per square meter depending on quantity and materials used. The North American agent for these cells is Electrosynthesis Co., Inc. (Lancaster, New York).

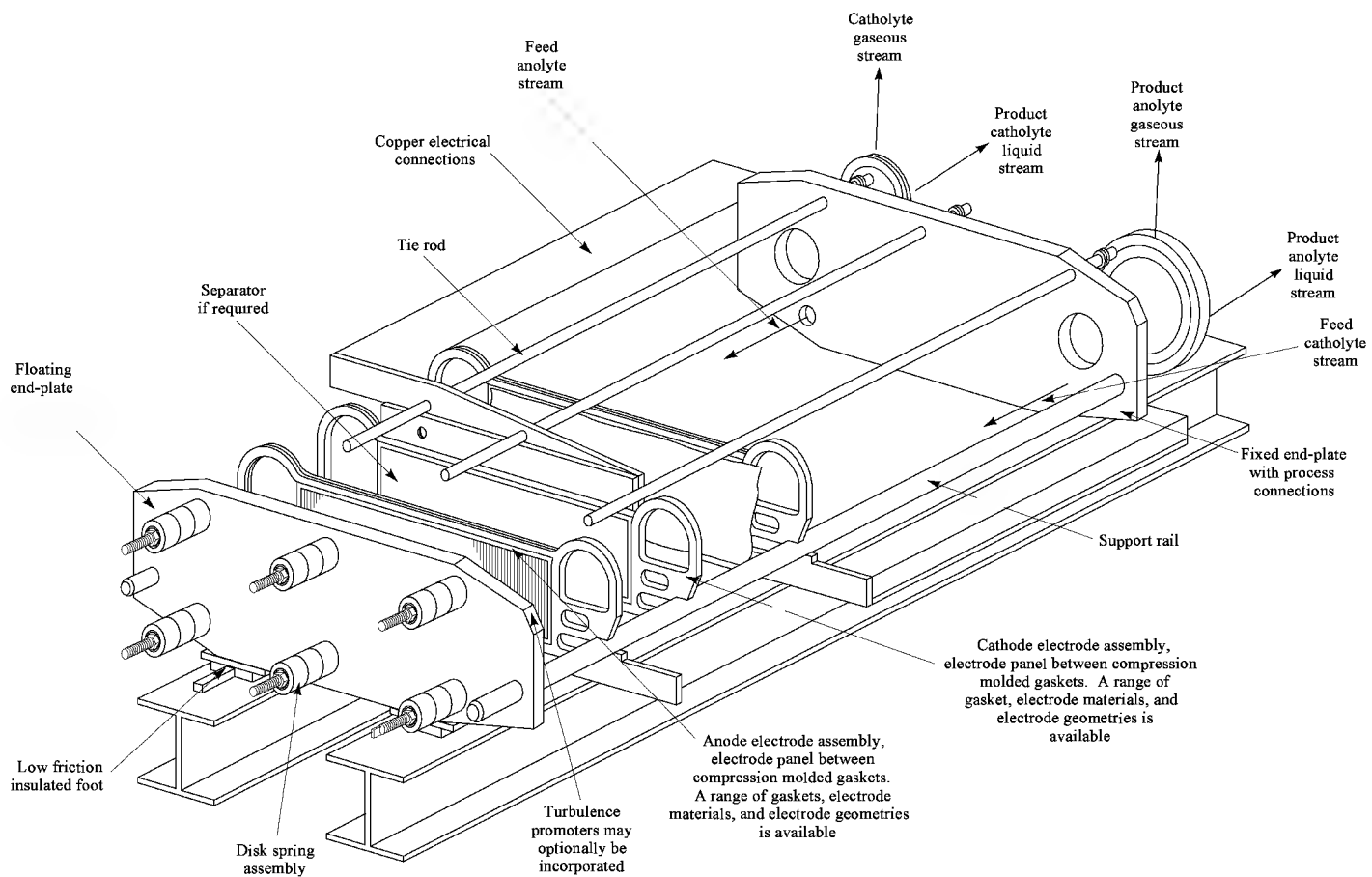


Fig. 11. The FM 21-SP cell of ICI.

Table 3. Features of the ICI Electrolysers^a

Cell	FM01-LC	FM 21-SP	FB01	FB20
filter press type	monopolar	monopolar	bipolar	bipolar
electrode area, m ²	0.0064	0.21	0.01	0.2
interelectrode gap, mm	1–10	1–10	1–10	1–10
electrolyte velocity per cell, m s	0–0.2	0–0.2	0–0.2	0–0.2

^a Electrode geometries may be plate, blade, or extended surface.

Electrocatalytic. The cell manufactured by Electrocatalytic (Union, New Jersey) was developed and patented (87) by the Electricity Council, Capenhurst, UK. The cell design is referred to as dished electrode membrane (DEM) cell, and is of the plate and frame type. The electrodes are dished to produce a wide entrance in the frame for the electrolyte supply. The design overcomes the problem of externally manifolding for a narrow gap parallel plate cell. This is illustrated in Figure 12 for both divided and undivided configurations. The electrode and diaphragm spacing is about 4 mm and turbulence promoters can be used, although high electrolyte velocity can be applied. Cell hardware comes in four electrode area sizes: 0.01, 0.05, 0.175, and 1.00 m². The external manifolding gives parallel electrolyte flow with a long electrolyte path to avoid current bypass. Electrical connection can be monopolar or bipolar. The current density is up to 500 mA cm², or in special cases up to 100 mA cm². Complete packages are available with pumps, tanks, piping and rectifier. Examples of this cell's use in electroorganic synthesis can be found in Reference 88. The cost of these cells is in the range of \$10,000–15,000 per square meter depending on quantity and materials used.

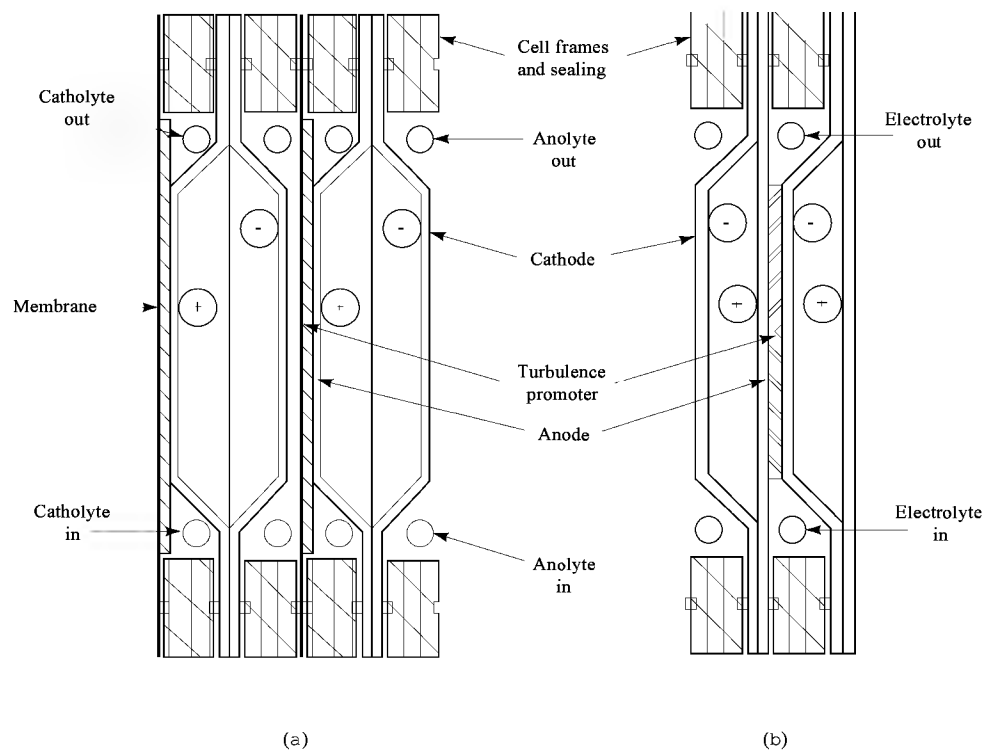


Fig. 12. Diagram of the DEM cell from Electrocatalytic, Inc. Cross sections of (a) divided, and (b), undivided configurations.

Aquanautics. Cells are also available from Aquanautics (Alameda, California) in three different sizes: a minicell having electrode dimensions of 5 × 5 cm; a bench cell having electrode area of 250 cm²; and a production cell 25 × 100 cm. The cells are designed to take advantage of high surface area electrode materials such as graphite felt. They can be configured for flow-by or flow-through of electrolyte. The cost of these cells is at the low end of the spectrum for commercial cells, ranging from \$5000–8000 per square meter.

Product Recovery. Comparison of the electrochemical cell to a chemical reactor shows the electrochemical cell to have two general features that impact product recovery. Cell product is usually liquid, can be aqueous, and is likely to contain electrolyte. In addition, there is a second product from the counter electrode, even if this is only a gas. Electrolyte conservation and purity are usual requirements. Because product separation from the starting material may be difficult, use of reaction to completion is desirable; cells would be run batch or plug flow. The water balance over the whole flow sheet needs to be considered, especially for divided cells where membranes transport a number of moles of water per Faraday. At the inception of a proposed electroorganic process, the product recovery and refining should be included in the evaluation to determine true viability. Thus early cell work needs to be carried out with the preferred electrolyte solvent and conversion. The economic aspects of product recovery strategies have been discussed (89). Some process flow sheets are also available (61).

Economic Aspects. Several publications probe the various areas of electroorganic process cost. Cells (90), overall process costs (41,91–93), economic optimization (94,95), and a comparison between the chemical and electrochemical methods (91,96) are all discussed.

Of the three key criteria that govern electrode processes, ie, product selectivity, electric power usage, and electrolysis system capital, any one of these may become paramount in the cost analysis. For instance, power cost is critical in a large-scale process such as the hydrodimerization of acrylonitrile. On the other hand, in small-scale high value added processes product selectivity is foremost. An important factor in the design of an electrolysis process is the intensity with which the cells can be driven, ie, how high a current density can be used. If it is assumed that the selectivity of the process is independent of current density, then capital and power costs are determining factors. The higher the current density, the less cell related capital is required. To balance this, the higher current density requires an increased power cost. When these two main factors are computed, a minimum total process operating cost is found at a particular current density. Thus the optimum capital and electrical energy usage point is found, and is used to size cells and rectifiers, etc.

Economic analyses including both a detailed cost estimate as well as a factored estimate of a generalized electroorganic process have been worked out (41). These estimates produce a similar result. A detailed breakdown of a cost estimate for a hypothetical electroorganic process using the factored method (41) is given in Table 4 for a process producing 10,000 t of product annually. The reaction is assumed to take two electrons per mole with a molecular weight of 100. In this analysis, only the cell related capital and operations are considered, no product extraction or refining. Electrical power, electrolyte cooling, rectifiers, pumps, tanks, and cells with membranes and electrodes, and maintenance and operating costs of the cell area are included. The cost of capital is taken to be 30% of the total investment. Results are illustrated in Figure 13 where it is seen that the cost of capital falls steeply with increasing current density, but tends to level out at the higher current densities. Electrical power costs increase continuously with current density because of the linear nature of cell voltage and current at well above reversible current densities. The membranes, maintenance, and labor factor fall with increasing current density, but are flatish for much of the range. The sum total conversion cost, excluding raw materials, is seen to go through a minimum at about 0.5 A cm². This is the optimum operating current density for this cost model.

Table 4. Current Density Optimization of a 10,000 t yr Organic Product^a

Factor	Models			
	1	2	3	4
cell current, A	1,050	2,100	4,200	8,400
current density, A cm ²	0.125	0.25	0.5	1.0
voltage per cell, V	4.9	5.8	7.5	11.0
number of cells	800	400	200	100
equipment cost, \$ × 10 ³				
cells ^b	3,200	1,600	800	400
rectifiers ^c	430	507	661	970
heat exchangers	332	367	430	542
pumps and tanks	448	244	136	78
total installed capital cost, \$ × 10 ³	17,775	10,722	7,686	7,123
operating cost, \$ × 10 ³ yr				
electricity ^d	1,340	1,580	2,060	3,020
cooling water	56	66	86	127

membranes ^e	360	180	90	45
cell maintenance	210	210	210	210
other maintenance	583	365	275	269
cell labor ^f	800	400	200	100
<i>total operating cost, \$ × 10³ yr</i>	<i>3,352</i>	<i>2,801</i>	<i>2,921</i>	<i>3,771</i>
capital charges at 30 ^o %	5,330	3,220	2,310	2,140
production cost, \$ × 10 ³ yr	8,682	6,021	5,231	5,911

^a Costs are estimated for 1992 and do not involve special materials or special cell design. Data were developed by the factor method (41) and do not purport to be any particular process. Equivalent weight of the reactant is 50, and 100^o % current efficiency is assumed.

^b Cost is \$4000 per cell, corresponding to a cell cost of \$4700 m² electrode area.

^c Cost is \$105,000 kW.

^d Cost is \$0.04 kWh.

^e Cost is \$450 per cell.

^f Cost is \$50,000 per operator.

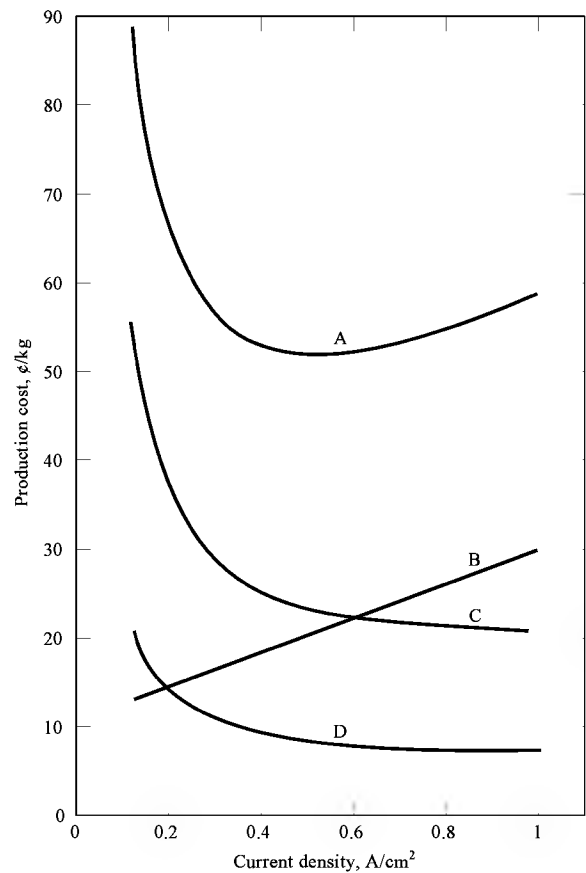


Fig. 13. Economic optimization of conversion costs for a plate and frame cell where A is total divided cell, B is electricity, C is capital, and D is membranes, maintenance, and labor.

For electrode processes that can be run in undivided cells, costs for the three factors are lowered. The cell voltage is lower, cells are less complicated and thus cheaper, and maintenance is simplified. Using a stack cell design, a new production cost for minimum current density is produced. This is shown in Figure 14 together with the total curve for the plate and frame cell. Here the undivided cell gives the lower cost and lower optimum current density of about 0.3 A cm². In this last example, if for some reason the cell had remained an expensive design but the power was still significantly lowered, then the current density minimum point would have been moved higher. Lowering the cost of capital related items moves the current density minimum down, and lowering the electrical power needs moves the current density minimum up.

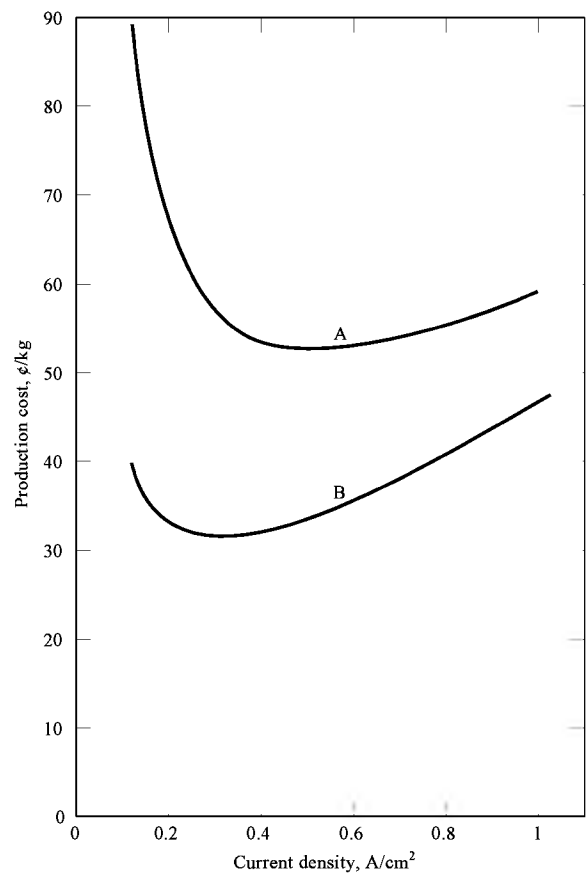


Fig. 14. Economic optimization comparison. Total cost of A, plate and frame, and B, stack cells.

Similarly, the optimum cost model for three cell designs including separation and refining costs has been examined (97). Using these cost models, it becomes clear where research and development is most advantageous. In most instances, there are clear advantages in fitting the process into narrow gap undivided cells if possible. Product separation and refining may dominate the cost model, as is the case of the sebacic acid [111-20-6] process. Here the demands of the cell reaction require the half ester which has to be recycled. Some general comments on product recovery have been made. Regarding the cost split between reaction area and the product work-up for electrochemical and conventional reactions (61), generally there is not much difference. However, in the scale-up of the cell reactor, the electrochemical method is at a disadvantage because of the near linear nature of cell scale-up. The cost model may require further complication for bipolar cells to take into account current leakage. Also, although product selectivity may remain constant, the current efficiency may change with current density, and should also be included in the cost model.

Industrial Electroorganic Processes

Some processing of organic chemicals by electrolysis for commercial purposes was carried out in the first half of the twentieth century. These were limited in number and mostly small-scale applications. The industrial uses of electroorganic synthesis on a large scale have been very few. The first large plant was built by Atlas Powder Co. in 1937 for the manufacture of sorbitol [50-70-4] and mannitol [87-78-5] using the cathodic reduction of glucose (see SUGAR ALCOHOLS) (98). Production was at the rate of 1400 t yr, but by 1947 this process was superseded by a high pressure catalytic hydrogenation method. In 1964, Nalco's tetraalkyllead process at Freeport, Texas came on stream. It had an original capacity of 14,000 t yr tetramethyllead [75-74-1], C₄H₁₂Pb, or 18,000 t yr tetraethyllead [78-00-2], C₈H₂₀Pb. In the next year, Monsanto started production of adiponitrile [111-69-3], C₆H₈N₂, by the electrohydrodimerization of acrylonitrile [107-13-1] at Decatur, Alabama, at similar rates.

Reports of bench-scale electroorganic reactions date back to the nineteenth century (18). Although the Nalco and Monsanto processes are the only two really large-scale operations, there has been significant growth in processes on a small scale. At least 60 processes appear to be commercial worldwide (4). A listing of some processes can be found in Table 5; further details and a listing of more processes are also available (4,5,25,88, and 99–104).

Table 5. Some Electroorganic Processes Operated on a Pilot-Plant or Commercial Scale

Product	Reactants	Company	Cell type ^a	Anode material	Cathode material
adiponitrile	acrylonitrile	Monsanto	P & F	Pb alloy	Pb
		Monsanto	St Ves	steel	Cd
		Asahi Chemical	P & F	Pb–Sb	Pb
		Asahi Chemical	P & F	Ni–steel	Pb
		BASF	St Ves	PbO ₂	C
		UCB-MCI	annular	magnetite	C
dimethyl sebacate	monomethyl adipate	Phillips	P & F	Pb	Pb
		Asahi Chemical	P & F	Pt–Ti	steel
		BASF	St Ves	Pt	C
		USSR	tank	Pt–Ti	stainless steel
tetraalkyllead	RCl, RMgCl, Pb	Nalco	packed bed	Pb	steel
propylene oxide	propylene	Kellog	tank	C	steel
		Bayer	tank	C	steel
acetylene dicar-boxylic acid	1,4-butyndiol	BASF	P & F	PbO ₂	Pb
1,2-dihydro-phthalic acid	o-phthalic acid	BASF	P & F	Pb	Pb
1,4-cyclohexadiene	benzene	Asahi Chemical	tank	Pt	Hg
		Mitsubishi Chemical	tank	Pt	Hg
		Monsanto	tank	steel	Hg

1,4-dihydro-naphthalene	naphthalene	Standard Oil	tank	C	Pb
dihydro-2-methoxy-naphthalene	2-methoxy-naphthalene	Hoechst	P & F	C	Hg–Cu
pinacol	acetone	BASF	tank	PbO ₂ –C	C
diacetone-2-ketogulonic acid	diacetone-L-sorbose	Diamond Shamrock	tank	DSA	Pb
ethylene glycol	formaldehyde	Hoffmann-LaRoche	swiss roll	Ni	Ni
maltol precursor	α-methyl furfural alcohol	Electrosynthesis Co.	P & F	Pt–Ti	C
3,4,5-trimethoxy toluene	3,4,5-trimethoxy tolyl	Otsuka Chemical	P & F	C	stainless steel
	alcohol	Otsuka Chemical	annular	C	stainless steel
4,4'-dihydroxy-hydrobenzoin	p-hydroxy-benzaldehyde	Monsanto	tank	steel	Hg

Table 5. (Continued)

Cell divider ^b	Temp, °C	Current density, A dm ²	Cell voltage, V	Product selectivity, %	Current efficiency, %	Power usage, kWh kg	Scale ^c	Reference
cation ex	55	20–100	6–12	90–92	90–92	6.6	C	80
none	55	20	4	87–89	80–85	2.3	C	80
cation ex	50	10–20	5–6	85–92	80–95	4	C	45
none	50	20	4	91	90	2.4	C	17
none	30	7–10	4–5	82–92	80–83	3	PP	74
none	20	4–15	5–6	72–92	84–93	3.7	PP	105
none	50	20	4	91	90	2.2	L	39
none	55	20	13–25	90–92	86–90	4	PP	106
none	40	25	12–15	80–85	60–70	5–6	PP	25
none	60	6	12	82	60	6	SC	107
polyethylene fabric	45	1.5–3	15–30	96–99	175	4–8	C	108
asbestos		80–110	3.5–4.0	82–86	70–80	5	PP	109
acrylic fabric	52	2–5	3–5	90	88	4–5	PP	110
cation ex	10	5		70	50		SC	25
cation ex	30–40	10	10	94	75	4.3	SC	25
cation ex	65	8		93	67		PP	111
cation ex				96	52		PP	112
none		2	5	90	99	3.4	PP	113
none	30–50	20–40	6–8	92	65	4–5	PP	114
cation ex	40–50	20	8	95	70	4	PP	115
none	20	10	7–10	90	20–30	15	SC	116
cation ex		16	5–6		70–80	4–5	PP	117
none	60	40	2.2	93	70	1.2	PP	118
cation ex	55	10	5–6	95	98	4–5	PP	119
none	5–10	10	7–15	92	97	3–5	PP	17
none	5–10	3–5	5–10	87	74	2–4	PP	17
none	45	8	5	95	75	1.6	PP	120

Table 5. (Continued)

Product	Reactants	Company	Cell type	Anode material	Cathode material
fenopropfen	CO ₂ + chloro-ethyldiphenyl-ether	SNPE	annular	Mg	stainless steel
tetramethyl-1,2,3,4,-butane-carboxylate	dimethylmaleate	Monsanto	modified SU	C	C
4-aminomethyl-pyridine	4-cyanopyridine	Reilly Industries	P & F	PbO ₂ –Pb	Pb
2-aminomethyl-pyridine	2-cyanopyridine	Reilly Industries	P & F	PbO ₂ –Pb	Pb
sorbitol	glucose	Atlas	tank	Pb	Pb–Hg
fluorinated organics	various organics	3M Phillips	tank tank	Ni C	Fe

Table 5. (Continued)

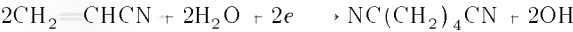
Cell divider ^b	Temp, °C	Current density, A dm ²	Cell voltage, V	Product selectivity, %	Current efficiency, %	Power usage, kWh kg	Scale ^c	Reference
none	30	5		80	66	6.6	PP	4
none	30	2.5–5	3–5	80	94	1.3	PP	121
cation ex		0.5–2	6.5	97	111	6.5	PP	100
cation ex		0.5–2	6.5	85	69	9.2	PP	100
ceramic		4	20		90	1.5–2	C	98
none		2	5–8				C	122

none	82	PP	16
^a P & F – plate and frame-type cell; St Ves – stacked electrodes placed in a closed vessel.			
^b Cation ex – cation-exchange membrane.			
^c C = commercial; L – laboratory; PP – pilot plant; and SC – semicommercial.			

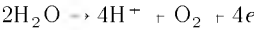
Large-Scale Processes.

Adiponitrile. The most significant commercial electroorganic synthesis process is Monsanto's electrohydrodimerization (EHD) of acrylonitrile to adiponitrile. The importance of adiponitrile is as a precursor to hexamethylenediamine [124-09-4] which is used in the manufacture of nylon-6, 6. Worldwide production of this polymer in 1987 was estimated at 1.8 million t yr, corresponding to 800,000 t yr adiponitrile. Table 6 gives the worldwide production figures for adiponitrile by the EHD process. The cost of manufacturing nylon-6,6 is critically dependent on the cost of the intermediates used, and this has maintained the pressure to produce improvements in the EHD process. In using concentrated solutions of quaternary ammonium salts (QAS), such as tetra-ethylammonium-*p*-toluenesufonate [733-44-8], C₈H₂₀N⁺·C₇H₇O₃S⁻, yields of about 90⁰ were demonstrated at lead and mercury cathodes (123). The principal reactions are as follows:

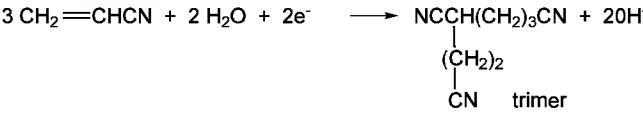
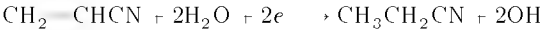
Main cathode reaction



Anodic water decomposition



Two principal cathode by-products



Propionitrile [107-12-0], C₃H₅N, and an acrylonitrile trimer, 1,3,6-tricyanohexane [1772-25-4] are also produced at the cathode.

Table 6. Worldwide EHD Plant Capacities

Company	Star-up year	Estimated capacity, t × 10 ³ yr
Monsanto, Decatur, Ala.	1965	118
BASF, Seal Sands, Teesside, UK ^a	1978	91
Asahi Chemical, Nobeoka, Japan	1971	37
Rhodia, Camaciri, Brazil ^b	1984	30

^a Former Monsanto and Montedison joint venture

^b Subsidiary of Rhône Poulenc.

Monsanto Cell/EHD Processes. The key cell parameters of Monsanto's divided and undivided cell processes are given in Table 7. The divided cell process is no longer used. The cell was of the plate and frame type (38) with external manifolding. Electrolyte flow was in parallel and recirculated through external heat exchangers; the anolyte was dilute sulfuric acid; the electrolysis system consisted of 16 cell presses, each containing 24 bipolar cells, and a current of 3000 A at 300 V was distributed to each cell bank. Four separate catholyte systems were used to feed the cell banks to minimize contamination in the event of a significant diaphragm rupture (38,80,124).

Table 7. Monsanto EHD Cell Configurations

Property	Divided cell ^a	Undivided cell ^b
cathode	Pb	Cd
anode	Pb ^c	steel
membrane	ionics CR 61	none
temp, °C	50	55
electrode gap, mm	7	1.8
electrolyte velocity, m s ⁻¹	2	1.2
current density, A dm ⁻²	45	20
cell voltage, V	11.7	3.8
power usage, kWh kg ⁻¹	6.6	2.3

^a The electrolyte contains 40 wt % (C₂H₅)₄NSO₄C₂H₅.

^b The electrolyte is comprised of 0.4 wt % hexamethylene bis(ethyldibutyl-ammonium) ion; 10 wt % Na₂HPO₄; 2 wt % Na₂B₄O₇; and 0.5 wt % tetrasodium ethylenediaminetetracetic acid.

^c Contains 1 wt % Ag.

In the divided cell EHD process, the catholyte was relatively expensive and a poor electrical conductor, the ion-exchange membrane was an added cost and gave added voltage drop, and the plate and frame cell was expensive to construct and maintain. These problems were addressed, and led to the development of an undivided cell technology (80). In the undivided cell, the electrolyte contains the good conducting phosphate ion and only low concentrations of the quaternary ammonium salt (QAS), which facilitates the dimerization. Sodium EDTA is present to solubilize the iron and cadmium electrode corrosion products and maintain cathode quality. Additionally, the cell structure was very much simplified. The membrane was eliminated, and the bipolar electrode structure was simplified by using cadmium-coated steel sheets. These rectangular sheets were simply stacked with insulator spacers in packs of 100–200 cells. The cell packs were then placed in an insulated vessel in a way to give uniform electrolyte flow (see Fig. 15) in the direction of the longer side.

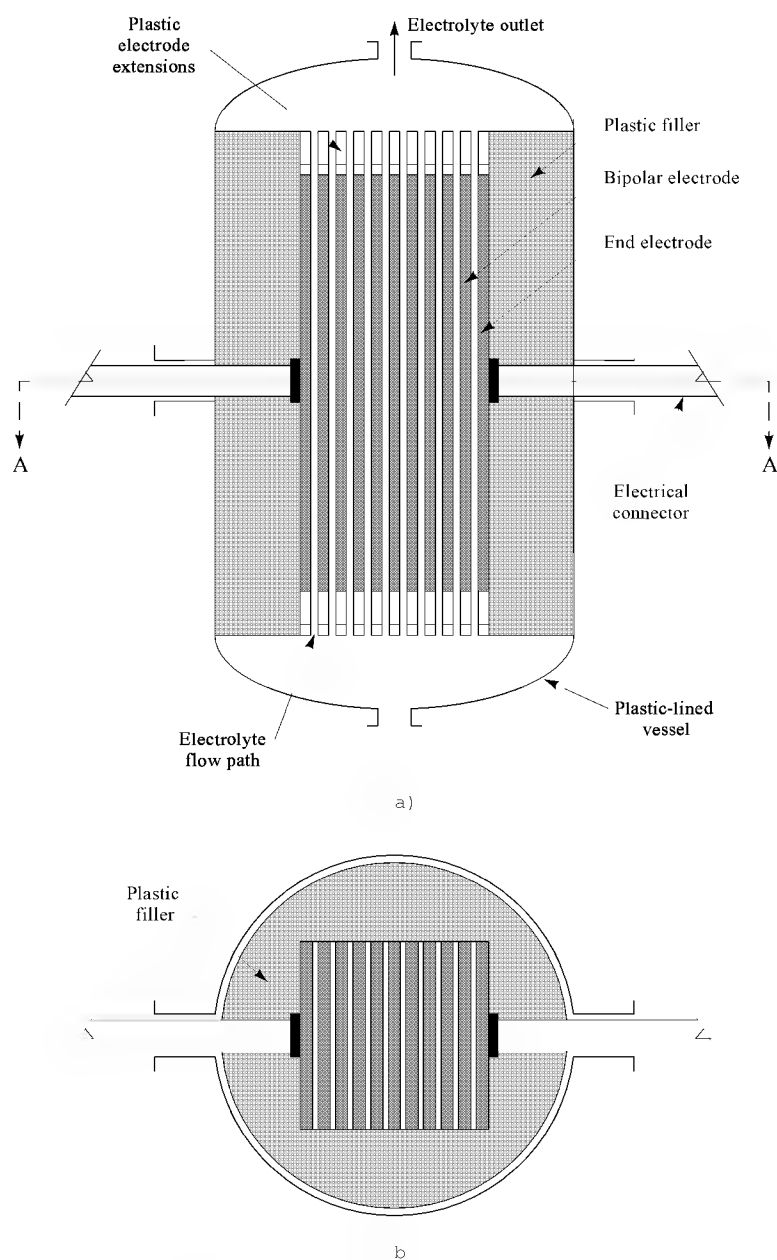


Fig. 15. Monsato EHD process stack cell (a) from side view; (b) section A–A (41).

To recover the adiponitrile product, the organic layer, containing acrylonitrile, adiponitrile, and by-products, is decanted from the electrolyte, and the QAS extracted from this using water. Figure 16 shows a flow diagram for this process. The water-containing QAS is returned to the cell operation. The organic stream leaving the extraction column is sent to a stripping column. There, following removal of by-product propionitrile, the acrylonitrile is recycled to the process. The tails from the acrylonitrile stripper are further refined by distillation to produce a 99 + % pure product. This product is suitable for hydrogenation into hexamethylenediamine. To remove the organic and metal ion impurities from the aqueous part of the electrolyte that can adversely affect the reaction selectivity and current efficiency, a portion of this is purged from the electrolyte circulating system. Adiponitrile is recovered from this stream by extraction with acrylonitrile. The aqueous part is then concentrated by evaporation and crystallized to recover phosphate and borate salts. The crystallizer mother liquor cadmium is effectively removed by precipitation.

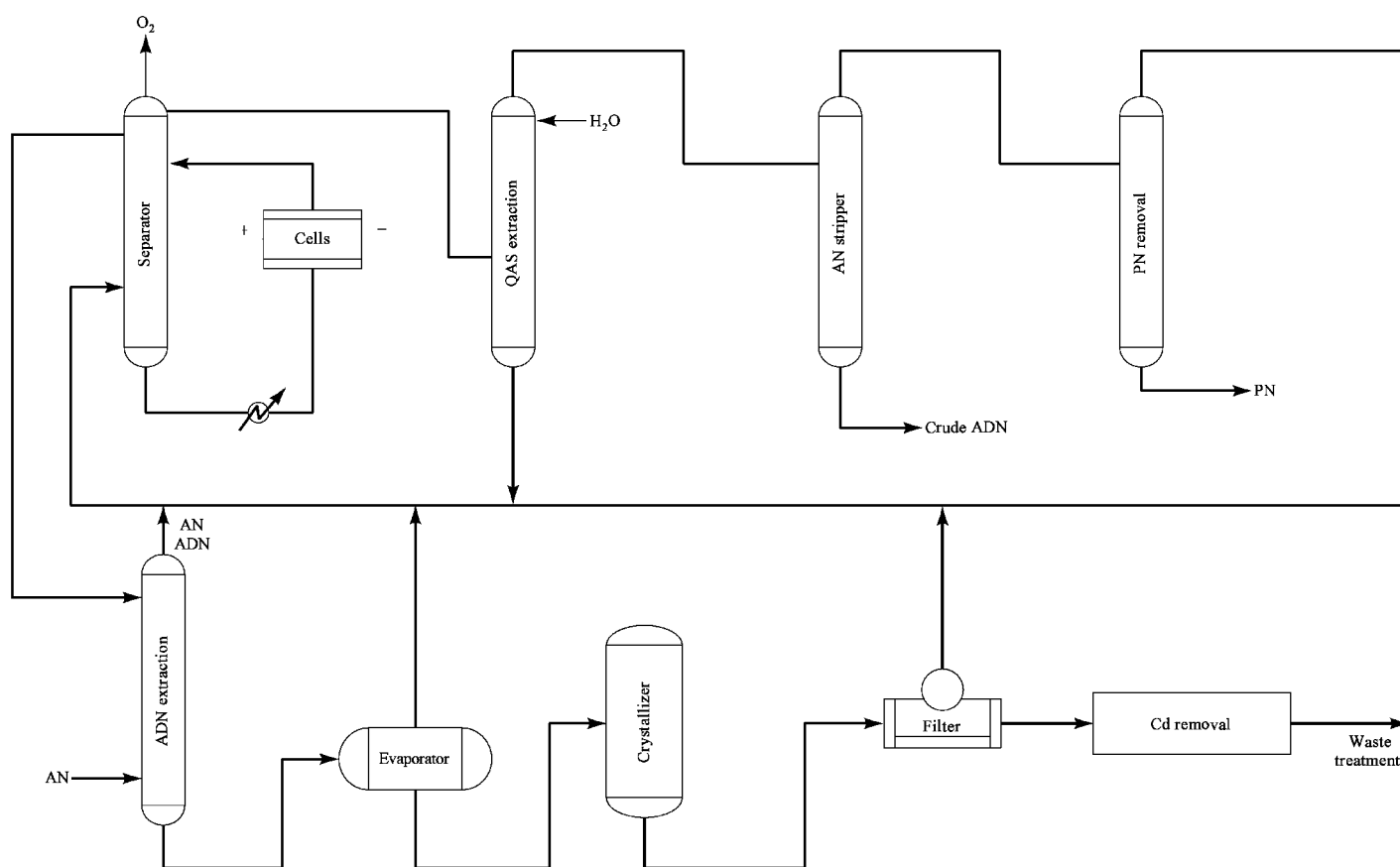


Fig. 16. Monsanto undivided cell EHD process flow sheet where AN is acrylonitrile; ADN adinonitrile; and PN polynitrile (41).

It is clear from Table 7 that the undivided cell has considerable power usage savings over the divided cell operation. Also, there are no membrane costs, and cell fabrication is much cheaper. In addition, it was possible to simplify the product recovery in the undivided cell process.

Asahi Chemical EHD Processes. In the late 1960s, Asahi Chemical Industries in Japan developed an alternative electrolyte system for the electroreductive coupling of acrylonitrile. The catholyte in the Asahi divided cell process consisted of an emulsion of acrylonitrile and electrolysis products in a 10% aqueous solution of tetraethylammonium sulfate. The concentration of acrylonitrile in the aqueous phase for the original Monsanto process was 15–20 wt %, but the Asahi process uses only about 2 wt %. Asahi claims simpler separation and purification of the adiponitrile from the catholyte. A cation-exchange membrane is employed with dilute sulfuric acid in the anode compartment. The cathode is lead containing 6% antimony, and the anode is the same alloy but also contains 0.7% silver (45). The current efficiency is of 88–89%, with an adiponitrile selectivity of 91%. This process, started by Asahi in 1971, at Nobeoka City, Japan, is also operated by the Rhône Poulenc subsidiary, Rhodia, in Brazil under license from Asahi.

Asahi also reports an undivided cell process employing a lead alloy cathode, a nickel–steel anode, and an electrolyte composed of an emulsion of 20 wt % of an oil phase and 80 wt % of an aqueous phase (125). The aqueous phase is 10 wt % K_2HPO_4 , 3 wt % $\text{K}_2\text{B}_4\text{O}_7$, and 2 wt % $(\text{C}_2\text{H}_5(\text{C}_4\text{H}_9)_3\text{N})_2\text{HPO}_4$. The oil phase is about 28 wt % acrylonitrile and 50 wt % adiponitrile. The balance of the oil phase consists of by-products and water. The cell operates at a current density of 20 A dm^{-2} at 50°C . Circulated across the cathode surface at a superficial velocity of 1.5 m s^{-1} is the electrolyte. A 91% selectivity to adiponitrile is claimed at a current efficiency of 90%. The respective anode and cathode corrosion rates are about mg (Ah) . Asahi's improved EHD process is reported to have been commercialized in 1987.

Tetraalkyllead.

The Nalco Process. The second greatest success story for electroorganic processing in terms of total tons of product is that of tetraalkyllead. However, because of the curtailment of the use of lead-based antiknock additives in gasoline, Nalco closed its commercial plant in the early 1980s. The Nalco process combined the large-scale production of Grignard reagent with a novel electrolysis system (108,126,127) (see also GRIGNARD REACTION). By using different alkyl groups on the Grignard reagent and the alkyl chloride fed to the cells, the process can be employed to prepare mixtures of tetraalkylleads.

Smaller-Scale Processes.

Fluorination. Perfluorinated organic compounds are important industrial surfactants (qv) and textile treating agents (see FLUORINE COMPOUNDS, ORGANIC; TEXTILES). Ammonium perfluorooctanoate [3825-26-1], $\text{C}_8\text{H}_4\text{F}_{15}\text{NO}_2$, has been employed as an emulsifying agent in the polymerization of tetrafluoroethylene (128). Other perfluorocompounds have been used to prepare a firefighting foam that has proven effective in smothering petroleum fires (129). A polymer formed from dihydropolyfluorooctyl acrylate has been reported to provide dirt repellency in a formulation applied to textile fabrics (130).

The discovery that a number of organic compounds could be fluorinated by electrolysis in a solution of anhydrous hydrogen fluoride [7664-39-3], HF, was made around 1940 (122) (see FLUORINE COMPOUNDS, INORGANIC HYDROGEN FLUORIDE). In the Simons process, fluorination of the dissolved organics takes place at a nickel anode, generally without generation of free fluorine; hydrogen is evolved at the iron cathode of diaphragmless cell. Alkali fluorides may be added for improved electrolyte conductivity. Typically, the electrolysis is carried out at low ($0\text{--}20^\circ\text{C}$) temperatures, low ($1\text{--}2 \text{ A dm}^{-2}$) current densities, and operating cell voltages of 5–6 V. Generally, a large number of products are formed, and the yields and current efficiencies are poor by commercial standards. Electrochemical fluorinations of a wide range of compounds including hydrocarbons, alcohols, ketones, carboxylic acids, and amines have been studied (131). A comprehensive review of the field has been published (132).

The first commercial production of electrofluorinated organics was carried out by 3M Co. at Hastings, Minnesota, in 1951. This facility employed a 10 kA cell and had a capacity of about 115 kg d of fluorocarbon products. The cell employed alternating planar nickel and iron sheets at 12.5 mm spacing, immersed in a cylindrical steel vessel 180 cm high by 120 cm diameter. Cooling was generally provided by coils of tubing within the cell vessel. Vapors consisting of hydrogen, hydrogen fluoride, and light fluorocarbons were passed through a condenser to remove the HF for return to the cell. This cell was used to produce a wide variety of fluorochemicals.

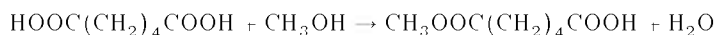
The electrofluorinated organics of principal industrial importance are perfluorooctanoic acid [335-67-1], $\text{CF}_3(\text{CF}_2)_7\text{COOH}$, and perfluorooctanesulfonic acid [1763-23-1], $\text{CF}_3(\text{CF}_2)_7\text{SO}_3\text{H}$. Typical conditions for preparation of the latter compound involve batch electrolysis of a 10 wt

° solution of *n*-octanesulfonyl chloride in anhydrous liquid hydrogen fluoride at 101 kPa (1 atm) and 17–19°C, employing an anode current density of 2 A dm². An insoluble liquid phase is formed during electrolysis and is readily withdrawn from the bottom of the cell; distillation of this material gives perfluorooctanesulfonyl fluoride in 30–35° molar yields. The potassium salt may be formed by hydrolysis of the sulfonyl fluoride with KOH and the corresponding acid is prepared by distilling from a mixture of the salt in 100° H₂SO₄.

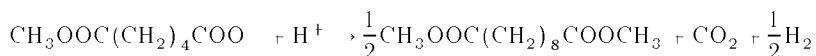
An improved electrofluorination system of Phillips Petroleum Co. has received extensive study. Instead of the flat nickel anode of the 3M cell, a porous carbon (qv) anode is employed, and the organic feedstock is introduced through the pores (132). By judicious control of the organic feed rate and anode current, varying degrees of fluorination can be achieved within the porous anode. Typically, a mixture of potassium fluoride and hydrogen fluoride (1:2 mole ratio) is employed as the electrolyte at operating temperatures of 60–105°C. Current densities of 6–30 A dm² are used, and cell voltages are generally 4–12 V. Current efficiencies to fluorinated products of 80–100° are reported using light hydrocarbon feedstocks. Whereas the 3M process prefers soluble organic reactants, the Phillips' process gives best results using insoluble organic reactant.

Sebacic Acid. Sebacic acid [111-20-6], C₁₀H₁₈O₄, is an important intermediate in the manufacture of polyamide resins (see POLYAMIDES). It has an estimated demand worldwide of approximately 20,000 t yr. The alkaline hydrolysis of castor oil (qv), which historically has shown some wide fluctuations in price, is the conventional method of preparation. Because of these price fluctuations, there have been years of considerable interest in an electrochemical route to sebacic acid based on adipic acid [124-04-9] (qv) as the starting material. The electrochemical step involves the Kolbŕ-type or Brown-Walker reaction where anodic coupling of the monomethyl ester of adipic acid forms dimethyl sebacate [106-79-6]. The three steps in the reaction sequence from adipic acid to sebacic acid are as follows:

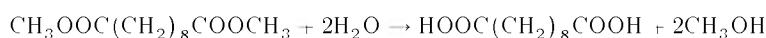
Formation of half-ester of adipic acid



Kolbŕ coupling of half-ester



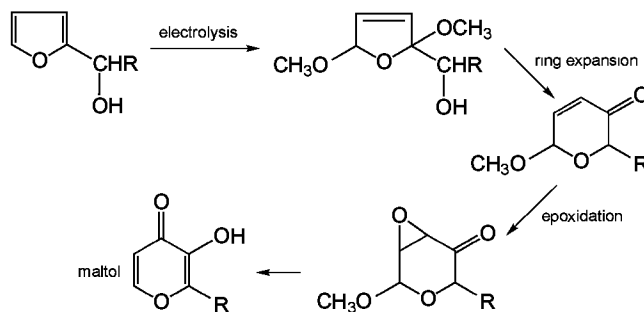
Hydrolysis of diester



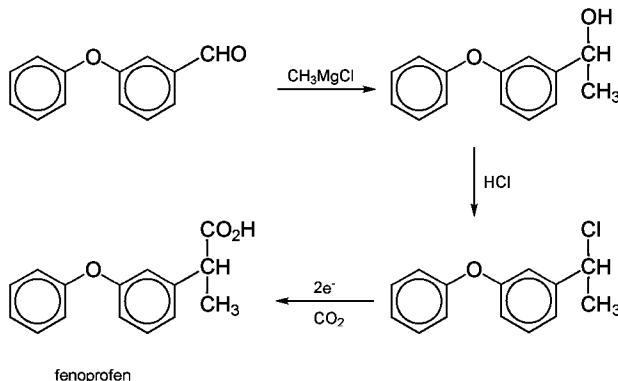
BASF (76), Asahi Chemical Industry (106), and a CIS group (107) have carried out pilot-plant studies of the electrochemical route to sebacic acid. A comparison of the electrolysis conditions employed by each may be found in Table 5. Yields of dimethyl sebacate reported for the BASF and CIS processes are in the range of 80–85°. Asahi claims product yields as high as 92° at current efficiencies in the range of 85–90°. The use of high electrolyte velocities (3–4 m s) through the cell is evidently the key to achieving these superior results. A 20 wt % aqueous solution of monomethyl adipate [627-91-8], having 30 mole per 100 mole of the organic acidity neutralized with sodium hydroxide, comprises the electrolyte and cell temperature is 55°C. A bipolar plate and frame cell with a platinum-plated titanium anode and a steel cathode is used in the Asahi process. Typically, the cells are operated at a current density of 20 A dm² and 14 V. The electrode spacing is 2 mm.

A flow diagram for the Asahi manufacture of sebacic acid from adipic acid is given in Reference 106. From the complexity of the process flow diagram, it appears that the sebacic acid process has significant investment and operating costs to overcome. The 1976 economic study by Chem Systems showed a slightly better than 10° cost advantage of the electrochemical route over the conventional process from castor oil in an 11,000 t yr plant. This technology was demonstrated by Asahi at a 100 t yr scale. As of this writing, Asahi has not commercialized the process. It is reported that sebacic acid has been produced by the electrochemical approach on a commercial scale in the former USSR.

Maltol. Otsuka Chemical Co. in Japan has operated several electroorganic processes on a small commercial scale. It has used plate and frame and annular cells at currents in the range of 4500–6000 A (133). The process for the synthesis of maltol [118-71-8], a food additive and flavor enhancer, starts from furfural [98-01-1] (see FOOD ADDITIVES; FLAVORS AND SPICES). The electrochemical step is the oxidation of α -methylfurfural to give a cyclic acetal. The remaining reaction sequence is acid-catalyzed ring expansion, epoxidation with hydrogen peroxide, and then acid-catalyzed rearrangement to yield maltol, i.e:

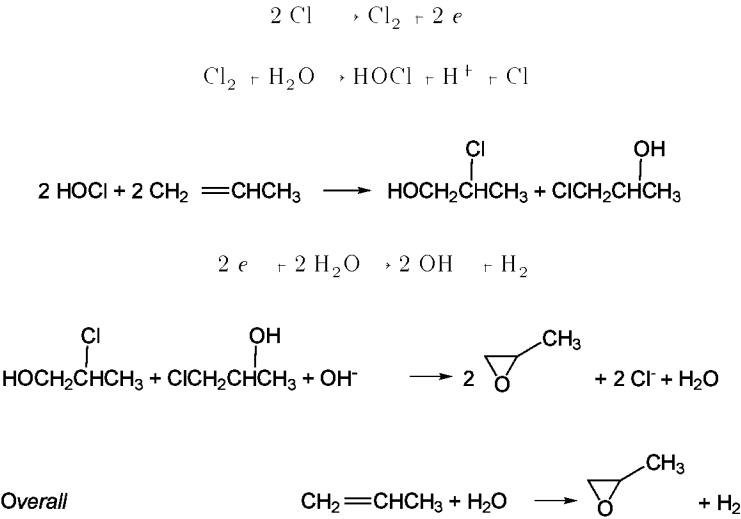


Fenoprofen. Fenoprofen [31879-50-7] is one of a series of antiinflammatory drugs of the arylpropionic acid family (see ANALGESICS, ANTIPYRETIC, AND ANTIINFLAMMATORY AGENTS). As is common with many complex drugs, its synthesis involves many steps of sometimes difficult chemistry. France's Soci  t   Nationale des Poudres et Explosifs (SNPE) uses an electrochemical step to replace a seven-step process by one of four steps (134). This is made possible by the availability of 3-phenoxybenzaldehyde, which is used as the starting point for the synthesis. The electrochemical step is an electrocarboxylation of the organic halide using carbon dioxide at a stainless steel cathode having some lead coverage. An aprotic solvent, such as dimethylformamide made conductive with a small amount of tetrabutylammonium halide is used, and the anode process uses the dissolution of magnesium. SNPE's reaction sequence is



In order to pilot this process, SNPE has developed an annular cell having a cone-shaped electrode section. The magnesium anode is placed into the conic section and fed like a pencil into a sharpener as it is consumed during electrolysis. The carbon dioxide is introduced into the electrolyte externally to the cell.

Propylene Oxide. An electrochemical route to 1,2-propylene oxide [75-56-9] has received considerable attention, but as yet has not been commercialized. The chemistry is related to one of the early processes for producing propylene oxide (qv) involving the reaction of propylene with chlorine and water to give the chlorohydrin followed by saponification with lime to form propylene oxide and by-product calcium chloride (see CHLOROHYDRINS). The electrochemical route involves passing gaseous propylene through the anode compartment of a divided cell in which both anolyte and catholyte are a dilute solution of sodium chloride. Hypochlorous acid produced in the anode compartment combines with the propylene to form propylene chlorohydrin, which diffuses through the porous diaphragm and is decomposed in the alkaline environment generated by the cathodic reduction of water. The principal reactions are



There have been a number of cell designs tested for this reaction. Undivided cells using sodium bromide electrolyte have been tried (see, for example, Ref. 29). These have had electrode shapes for in-cell propylene absorption into the electrolyte. The chief advantages of the electrochemical route to propylene oxide are elimination of the need for chlorine and lime, as well as avoidance of calcium chloride disposal (see CALCIUM COMPOUNDS, CALCIUM CHLORIDE; LIME AND LIMESTONE). An indirect electrochemical approach meeting these same objectives employs the chlorine produced at the anode of a membrane cell for preparing the propylene chlorohydrin external to the electrolysis system. The caustic made at the cathode is used to convert the chlorohydrin to propylene oxide, reforming a NaCl solution which is recycled. Attractive economics are claimed for this combined chlor-alkali electrolysis and propylene oxide manufacture (135).

BIBLIOGRAPHY

"Electrochemistry" in *ECT* 1st ed., Vol. 5, pp. 495–549, by G. C. Akerlog, Mellon Institute of Industrial Research; in *ECT* 2nd ed., Vol. 7, pp. 784–841, by G. C. Akerlog, Princeton University; "Organic" under "Electrochemical Processing" in *ECT* 3rd ed., Vol. 8, pp. 696–720, by D. Danly, Monsanto Chemical Intermediates Co.

1. J. R. Selman in R. C. Alkire and T. Beck, eds., *Tutorial Lectures in Electrochemical Engineering and Technology*, AIChE Symposium Series, Vol. 77, number 204, AIChE, New York, 1981, p. 88.
2. G. Prentice, *Electrochemical Engineering Principles*, Prentice Hall, Inc., Englewood Cliffs, N.J., 1991.
3. M. I. Ismail, ed., *Electrochemical Reactors, Their Science and Technology*, Part A, Elsevier Science Publishing Co., Inc., New York, 1989.
4. J. D. Genders and D. Pletcher, eds., *Electrosynthesis, From Laboratory, to Pilot, to Production*, Electrosynthesis Co., E. Amherst, N.Y., 1990.
5. D. Pletcher and F. C. Walsh, *Industrial Electrochemistry*, 2nd ed., Chapman and Hall, New York, 1990.
6. K. Scott, *Electrochemical Reaction Engineering*, Academic Press, Inc., New York, 1991.
7. U. Landau, E. Yeager, and D. Kortan, eds., *Electrochemistry in Industry*, Plenum Press, New York, 1982.
8. D. J. Pickett, *Electrochemical Reactor Design*, Elsevier Science Publishing Co., Inc., New York, 1977.
9. T. Z. Fahidy, *Principles of Electrochemical Reactor Analysis*, Elsevier Science Publishing Co., Inc., New York, 1985.
10. F. Coeuret and A. Storck, *Elements de Genie Electrochimique*, Lavoisier, Paris, 1984.
11. J. S. Newman, *Electrochemical Systems*, 2nd ed., Prentice Hall, Inc., New York, 1991.
12. E. Heitz and G. Kreysa, *Principles of Electrochemical Engineering*, VCH Publishers, New York, 1986.
13. L. Rousar, K. Micka, and A. Kimla, *Electrochemical Engineering*, Vols. 1 and 2, Elsevier Science Publishing Co., Inc., New York, 1986.
14. F. Hine, *Electrode Processes and Electrochemical Engineering*, Plenum Press, New York, 1985.
15. R. D. Little and N. L. Weinberg, eds., *Electroorganic Synthesis, Festschrift for Manuel M. Baizer*, Marcel Dekker, Inc., New York, 1991.
16. M. M. Baizer and H. Lund, eds., *Organic Electrochemistry*, Marcel Dekker, Inc., New York, 1991.
17. S. Torii, ed., *Recent Advances in Electroorganic Synthesis*, Studies in Organic Chemistry 30, Elsevier Science Publishing Co., Inc., New York, 1987.
18. R. C. Alkire and S. Swan, Jr., *Bibliography of Electroorganic Synthesis, 1801–1975*, The Electrochemical Society, Princeton, 1980.
19. A. J. Fry, *Synthetic Organic Electrochemistry*, 2nd ed., John Wiley & Sons, Inc., New York, 1989.
20. M. R. Rifi and F. H. Covitz, *Introduction to Organic Electrochemistry*, Marcel Dekker, Inc., New York, 1974.
21. F. Beck, *Electroorganische Chemie*, Verlag Chemie, Weinheim, 1974.
22. A. P. Tomolov and co-workers, *Electrochemistry of Organic Compounds*, transl. J. Schmorak, Halsted Press, New York, 1972.
23. T. Shono, *Electroorganic Chemistry as a New Tool in Organic Synthesis*, Springer-Verlag, New York, 1984.
24. N. L. Weinberg, ed., *Technique of Electroorganic Synthesis*, Parts I and II, A. Weissberger, series ed., *Techniques of Chemistry*, Vol. 5, John Wiley & Sons, Inc., New York, 1974.
25. N. L. Weinberg and B. V. Tilak, eds., *Technique of Electroorganic Synthesis*, Part III, A. Weissberger, series ed., *Techniques of Chemistry*, Vol. 5, John Wiley & Sons, Inc., New York, 1982, Chapt. 4.
26. A. M. Couper, D. Pletcher, and F. C. Walsh, *Chem. Rev.* **90**, 837 (1990).
27. A. T. Kuhn, ed., *Industrial Electrochemical Processes*, Elsevier Science Publishing Co., Inc., New York, 1971, Chapt. 14.
28. J. P. Carr and N. A. Hampson, *Chem. Rev.* **72**(6), 679 (1972).
29. C. J. H. King, K. Lister, and R. E. Plimley, *Trans. Inst. Chem. Eng.* **53**(1), 20 (1975).
30. M. Hayes and A. T. Kuhn, *J. Appl. Electrochem.* **8**, 327 (1978).
31. D. C. Silverman, *Corrosion-NACE* **38**, 541 (1982).
32. P. M. Robertson, *Electrochimica Acta* **22**, 411 (1977).

33. R. B. MacMullin and G. A. Muccini, *J. Am. Inst. Chem. Eng.* **2**, 393 (1956).
34. G. Richter, *Chem.-Ing.-Tech.* **47**, 909 (1975).
35. W. Pusch, *Chem.-Ing.-Tech.* **47**, 914 (1975).
36. S. C. Stitson, *Chem. Eng. News*, 22 (Mar. 15, 1982).
37. Ref. 7, p. 73.
38. Ref. 25, Chapt. 6.
39. M. Krumpelt, E. Y. Weissman, and R. C. Alkire, eds., *Electro-organic Synthesis Technology*, AIChE Symposium Series, Vol. 75, number 185, AIChE, New York, 1979, p. 109.
40. Ref. 16, Chapt. 6.
41. D. E. Danly, *Emerging Opportunities for Electroorganic Processes*, Marcel Dekker, Inc., New York, 1984.
42. J. R. Selman and C. W. Tobias in T. B. Drew and co-workers, eds., *Advances in Chemical Engineering*, Vol. 10, Academic Press, Inc., New York, 1978, p. 211.
43. R. C. Alkire and T. Beck, eds., in Ref. 1, pp. 46 and 88.
44. D. E. Danly, *Chemistry and Industry*, (13), 439 (1979).
45. M. Seko, *Chem. Econ. Eng. Rev.* **7**(10), 20 (1975).
46. Ref. 25, Chapt. 8.
47. R. W. Houghton and A. T. Kuhn, *J. Appl. Electrochem.* **4**, 173 (1974).
48. Ref. 1, pp. 46 and 138.
49. Ref. 16, p. 1296.
50. C. W. Tobias, *J. Electrochem. Soc.* **106**, 833 (1959).
51. J. E. Funk and J. F. Thorpe, *J. Electrochem. Soc.* **116**, 48 (1969).
52. Y. Nisliki and co-workers, *J. Appl. Electrochem.* **16**, 615 (1986).
53. R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena*, John Wiley & Sons, Inc., New York, 1960.
54. G. Rogers and Y. Mayhew, *Engineering Thermodynamics, Work and Heat Transfer*, 3rd ed., Longman, London, 1980.
55. Ref. 16, p. 1300.
56. F. Lapique and A. Storck, *J. Appl. Electrochem.* **15**, 925 (1985).
57. Ref. 16, p. 1301.
58. T. V. Nyuyen, C. W. Walton, and R. E. White, *J. Electrochem. Soc.* **133**, 1130 (1986).
59. S. V. Ho, *J. Appl. Electrochem.* **17**, 398 (1987).
60. Ref. 39, p. 19.
61. R. C. Alkire and D-T. Chin, eds., *Tutorial Lectures in Electrochemical Engineering and Technology-II*, AIChE Symposium Series, Vol. 79, number 229, AIChE, New York, 1983, p. 79.
62. R. E. W. Jansson, *Chem. Eng. Sci.* **35**, 1979 (1980).
63. Ref. 61, p. 101.
64. R. B. MacMullin, *Electrochem. Technol.* **1**, 5 (1963).
65. *Ibid.*, **2**, 106 (1964).
66. Ref. 16, p. 1302.
67. Ref. 16, p. 1303.
68. Ref. 39, p. 8.
69. R. E. White and co-workers, *J. Electrochem. Soc.* **133**, 485 (1986).
70. Ref. 61, p. 68.
71. R. J. Marshall and F. C. Walsh, *Surface Technol.* **24**, 45 (1985).
72. U.S. Pat. **3,119,760** (Jan. 28, 1964), R. W. Foreman and F. Veatch (to Standard Oil of Ohio).
73. U.S. Pat. **3,119,770** (1964).
74. F. Beck and H. Guthke, *Chem.-Ing.-Tech.* **41**, 943 (1969).
75. F. Beck, *J. Appl. Electrochem.* **2**, 59 (1972).
76. Ref. 25, Chapt. 5.
77. U.S. Pat. **4,406,768** (Sept. 27, 1983), C. J. H. King (to Monsanto).
78. U.S. Pat. **4,406,768** (1983).
79. U.S. Pat. **4,500,403** (Feb. 19, 1985), C. J. H. King (to Monsanto).
80. D. E. Danly, *J. Electrochem. Soc.* **131**, 435C (1984).
81. Ref. 25, p. 311.
82. L. L. Bott, *Hydrocarbon Proc.* **44**(1), 115 (1965).
83. L. Carlsson, B. Sandegren, and D. Simonsson, *J. Electrochem. Soc.* **130**, 342 (1983).
84. Ref. 25, Chapt. 3.
85. Ref. 17, pp. 445 and 449.
86. W. N. Brooks, *I. Ch. E. Symp. Series*, (98), 1 (1986).
87. Brit. Pat. **2,098,238** (1984), J. P. Millington and I. M. Dalrymple (to Electricity Council).
88. D. Blum and P. Millington, *Chemicals*, 469 (Nov. 1990).
89. R. E. W. Jansson, *DECHEMA Monogr.* **112**, 211 (1988).
90. Ref. 39, p. 76.
91. Ref. 25, Chapt. 9 and 10.
92. R. E. W. Jansson, *J. Appl. Electrochem.* **12**, 163 (1982).
93. R. M. Spotnitz, R. W. Spillman, and J. T. Lundquist, *J. Electrochem. Soc.* **134**, 2435 (1987).
94. A. R. Despic, *Bull. Soc. Chim. (Belgrade)* **42**, 693 (1977).
95. R. Keller, *Electrochimica Acta* **25**, 303 (1980).
96. J. L. Fitzjohn, *Chem. Eng. Prog.* **71**(2), 85 (1975).
97. Ref. 61, p. 92.
98. R. L. Taylor, *Chem. Met. Eng.* **43**, 588 (1937).
99. D. Degner in E. Steckhan, ed., *Organic Electrosynthesis in Industry, Electrochemistry III, Topics in Current Chemistry*, Vol. 148, Springer-Verlag, Berlin, 1988, p. 1.
100. J. E. Toomey, and J. C. Yu, *Chem. Eng.*, 140 (June 1989).
101. T. R. Beck and co-workers, *A Survey of Organic Electrolytic Processes*, U.S. Department of Commerce, National Technical Information Service, Report No. ANL OEPM-79-5, Washington, D.C., Nov. 1979.
102. G. M. Graff, *Chem. Eng.*, 14 (Mar. 21, 1983).
103. G. Samdani and K. Gilges, *Chem. Eng.*, 37 (May 1991).
104. J. E. Toomey, Jr., P. B. Lawin, and G. A. Chaney, *Chemtech*, 37 (Jan. 1992).
105. C. Van Eygen and co-workers, *Chem. Ind.* **104**(1), 71 (1971).

106. M. Seko, A. Yomiyama, and T. Isoya, *Hydrocarbon Proc.*, 117 (Dec. 1979).
107. Yu.B. Vassiliev, E. P. Kovsman, and G. N. Friedlin, *Electrochimica Acta* **27**(7), 953 (1982).
108. L. L. Bott, *Hydrocarbon Proc.* **44**(1), 115 (1965).
109. **U.S. Pat. 3,342,717** (Sept. 19, 1967), J. A. M. Le Duc (to Pullman, Inc.).
110. **Brit. Pat. 1,090,006** (Nov. 8, 1967), W. Kronig and J. Grolig (to Farbenfabriken Bayer AG).
111. **U.S. Pat. 3,700,572** (Oct. 24, 1972), T. Hatayama, Y. Hamano, and T. Yamamoto (to Asahi Kasei).
112. **U.S. Pat. 3,485,726** (Dec. 23, 1969), A. Minono and T Osa (to Mitsubishi Chemical Industries Ltd.).
113. J. P. Coleman and J. H. Wagenknecht, *J. Electrochem. Soc.* **128**, 322 (1981).
114. **U.S. Pat. 4,022,673** (May 10, 1977), J. F. Connolly (to Standard Oil Co.).
115. **U.S. Pat. 4,115,216** (Sept. 19, 1978), D. H. Skaletz (to Hoechst AG).
116. **U.S. Pat. 3,899,401** (Aug. 12, 1975), H. Nohe and F. Beck (to BASF AG).
117. **U.S. Pat. 3,992,269** (Nov. 16, 1976), T. T. Sugano and co-workers, (to Diamond Shamrock Corp.).
118. P. M. Robertson and co-workers, *J. Electrochem. Soc.* **130**, 593 (1983).
119. **U.S. Pat. 4,478,694** (Oct. 23, 1984), N. L. Weinberg (to SKA Associates).
120. C. J. H. King, *J. Chem. Tech. Biotechnol.* **29**, 741 (1979).
121. **Eur. Pat. Appl. 433,260** (June 19, 1991), J. H. Wagenknecht, E. A. Casanova, and D. J. Kalota (to Monsanto).
122. J. H. Simons in J. H. Simons, ed., *Fluorine Chemistry*, Vol. I, Academic Press, Inc., New York, 1950.
123. M. M. Baizer, *Chemtech* **10**(3), 161 (1980).
124. D. E. Danly, *Chemtech* **10**(5), 302 (1980).
125. Ref. 17, p. 415.
126. **U.S. Pat. 3,391,067** (July 2, 1968), D. G. Braithwaite (to Nalco Chemical Co.).
127. **U.S. Pat. 3,372,097** (Mar. 5, 1968), D. G. Braithwaite and L. L. Bott (to Nalco Chemical Co.)
128. **Brit. Pat. 1,071,992** (June 14, 1967), D. L. Schindler (to E. I. du Pont de Nemours & Co., Inc.).
129. **U.S. Pat. 3,258,423** (June 28, 1966), R. L. Tuve and E. J. Jablonski (to Secretary of U.S. Navy).
130. E. L. Grajeck and W. H. Peterson, *Text. Res. J.* **32**, 320 (1962).
131. J. H. Simons and co-workers, *J. Electrochem. Soc.* **95**(2), 47 (1949).
132. Ref. 16, Chapt. 26.
133. Ref. 17, p. 433.
134. Ref. 4, Chapt. 8.
135. Ref. 39. p. 89.

Chris J. H. King
Monsanto Chemical Group

ELECTROCHEMISTRY.

See BATTERIES; ELECTROANALYTICAL TECHNIQUES; ELECTROCHEMICAL PROCESSING; ELECTROSEPARATIONS.

ELECTROLESS PLATING

The formation of a continuous metallic film on a surface without the use of an electric current and when done by a definite process is called electroless plating (1). Electroless plating is defined as the controlled autocatalytic deposition of a continuous film at a catalytic interface by the reaction in solution of a metal salt and a chemical reducing agent. This definition excludes such processes as conventional electroplating (qv), which uses externally supplied electrons as the reducing agent; uncontrolled precipitation of metal powders; immersion films produced by displacement reactions, often mistakenly labeled as electroless processes; thermal decomposition films; and the various types of sputtering and evaporation techniques (see THIN FILMS). Electroless deposition can produce films of metals, alloys, compounds, and composites on both conductive and nonconductive surfaces (2–5) (see also METALLIC COATINGS).

The first description of electroless plating was that of von Liebig in 1835 involving the reduction of silver salts by reducing aldehydes (qv). Progress in this field remained slow, however, until World War II (6). One significant problem was the need for very high purity chemicals, as the surfaces being plated must maintain autocatalytic activity. Many substances react with and poison the catalytic surface even if present only in trace amounts. Impurities can also cause the opposite problem, overactivity. Colloidal materials (see COLLOIDS) and fine particles can promote uncontrolled autocatalytic reactions owing to extremely high surface areas. Another reason for the lack of wide interest in electroless plating was cost. Chemical reducing agents are much more expensive than electricity.

The growth of electroless plating is directly traceable to: (1) the discovery that some alloys produced by electroless deposition, notably nickel phosphorus, have unique properties; (2) the growth of the electronics industry, especially the development of printed circuits (see ELECTRONIC COATINGS; INTEGRATED CIRCUITS); and (3) the large-scale introduction of plastics into everyday life.

Modern electroless plating began in 1944 with the rediscovery that hypophosphite could bring about nickel deposition (7,8). Subsequent work led to the first patents on commercially usable electroless nickel solutions. Although these solutions were very useful for coating metals, they could not be used on most plastics because the operating temperature was 90–100°C. The first electroless nickel solution capable of wide use on plastics was introduced in 1966 (9). This solution was usable at room temperature and was extremely stable (see NICKEL AND NICKEL ALLOYS).

Electroless copper solutions underwent similar development during the same period (10). Early printed circuit boards used mechanically attached eyelets to provide electrical conductivity between the copper sheathing laminated on two sides of a plastic board. Electroless copper plating provided a less expensive, better conductive path, allowing much greater numbers and smaller sizes of holes. Later, electroless coppers even replaced the laminated bulk copper sheathing in the semiadditive and additive processes (see COPPER).

Interest in electroless plating grew as it was shown that uniformity of deposition was very high. Current density complications, customized anodes, and special plating fixtures needed for the use of electrolytic plating were eliminated. The lack of requirements for electrical continuity made electroless baths an ideal solution to the problem of plating unconnected metal areas on printed circuit boards. In 1989 the total U.S. market for electroless plating chemicals was estimated to be \$347 million, and the projected growth was to \$525 million by 1994 (11).

Theory of Electroless Plating

The theory and practice of electroless plating parallels that of electrolytic plating.

The actual metal reduction and film development occurs at the interface of the solution and the item being plated in both the electrolytic and electroless processes. The main difference is that the electrons in electroless plating are supplied by a chemical reducing agent present in solution. In electroplating, these electrons are supplied by an external source such as a battery or generator. This means that electroless solutions are not thermodynamically stable, because the reducing agent and the metal salt are always present and ready to react.

The simplest electroplating baths consist of a solution of a soluble metal salt. Electrons are supplied to the conductive metal surface, where electron transfer to and reduction of the dissolved metal ions occur. Such simple electroplating baths are rarely satisfactory, and additives are required to control conductivity, pH, crystal structure, throwing power, and other conditions.

Electroless solutions contain a metal salt, a reducing agent, a pH adjuster or buffer, a complexing agent, and one or more additives to control stability, film properties, deposition rates, etc.

Of the large number of potential reducing agents, the principal commercial materials are formaldehyde (qv) [50-0-0] for copper and silver, hypophosphite for nickel and palladium, and organoboron compounds for gold, nickel, palladium, and copper. The latter two reducing agents produce phosphorus- or boron-containing alloys. The detailed theory of electroless plating has been discussed thoroughly in a number of works (3,10,12–16).

Most reducing agents are too slow, giving insufficient plating rates, or too fast, resulting in bulk decomposition. Each combination of metals and reducing agent requires a specific pH range and bath formulation. The metal salt and reducing agent must be replenished at periodic intervals. Buffers, complexers, and other additives are added to compensate for drag-out losses, and to counter bath aging effects such as total salts accumulation. Stabilizers are typically used at a few to a few tens ppm, and may affect deposition rate, deposit color, or metal ductility, or control spontaneous decomposition reactions.

The ideal electroless solution deposits metal only on an immersed article, never as a film on the sides of the tank or as a fine powder. Room temperature electroless nickel baths closely approach this ideal; electroless copper plating is beginning to approach this stability when carefully controlled. Any metal that can be electroplated can theoretically also be deposited by electroless plating. Only a few metals, ie, nickel, copper, gold, palladium, and silver, are used on any significant commercial scale.

Electrolytic plating rates are controlled by the current density at the metal–solution interface. The current distribution on a complex part is never uniform, and this can lead to large differences in plating rate and deposit thickness over the part surface. Uniform plating of blind holes, re-entrant cavities, and long projections is especially difficult.

A primary advantage of electroless solutions is the ability to produce conductive metallic films on properly prepared nonconductors, along with the ability to uniformly coat any platable object. The most complex geometric shapes receive a uniform plated film. Film thicknesses range from <0.1 μm, where only conductivity or reflectivity is wanted, to >1 mm for functional applications.

Electroless plating rates are affected by the rate of reduction of the dissolved reducing agent and the dissolved metal ion which diffuse to the catalytic surface of the object being plated. When an initial continuous metal film is deposited, the whole surface is at one potential determined by the mixed potential of the system (17). The current density is the same everywhere on the surface as long as flow and diffusion are unrestricted so the metal

deposited is uniform in thickness over the whole surface. However, maximum plating rates are usually lower for electroless plating than those possible for electrolytic plating. Extremely thin films of electroless coatings are not uniform, because the initial deposition is confined to discrete nucleation sites that grow and coalesce into a film. The difference between very thin electroless silver and sputtered silver is shown in Figure 1.

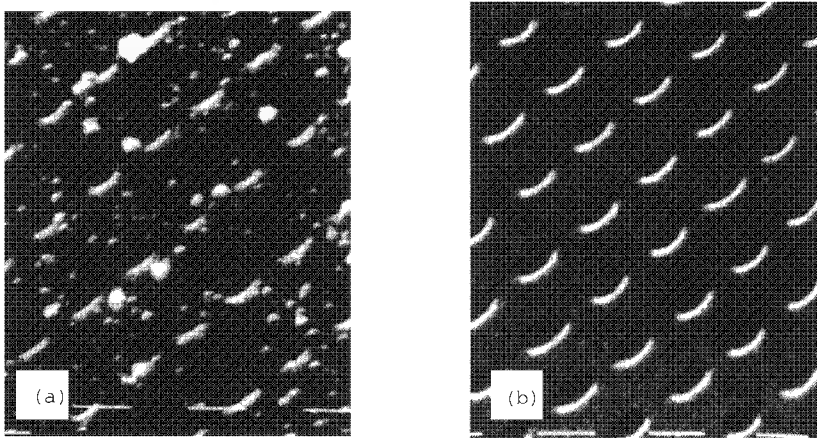


Fig. 1. Lacquer-coated optical readout laser disk master. Plating by (a) electroless silver spray coating and by (b) vacuum evaporation. Scale bar, 10⁻⁴ cm.

Courtesy of Nederlandse Philips Bedrijven BV.

Deposition reactions for some reducing agents are given in Table 1; hydrogen is a principal by-product of each reduction. Elemental phosphorus or boron is codeposited with the reduced metal from hypophosphite, borohydride, or organoborane baths (15). Other minor reactions can also occur (18). All of these reductions can be viewed as dehydrogenation reactions (16,19).

Table 1. Deposition Reactions

Reducing agent	Cathodic reactions ^a	Anodic reactions	Bath pH
formaldehyde ^{b,c}	$M^{z+} + xe^- \rightarrow M^0$ $2\text{H}_2\text{O} + 2e^- \rightarrow 2\text{OH}^- + \text{H}_2$	$2\text{HCHO} + 4\text{OH}^- \rightarrow 2\text{HCOO}^- + 2\text{H}_2\text{O} + 2e^-$ $\text{HCHO} + 3\text{OH}^- \rightarrow \text{HCOO}^- + 2\text{H}_2\text{O} + 2e^-$	11–13
sodium hypophosphite ^d	$M^{z+} + xe^- \rightarrow M^0$ $\text{H}_2\text{PO}_2^- + e^- \rightarrow \text{P}^0 + 2\text{OH}^-$ $2\text{H}_2\text{O} + 2e^- \rightarrow 2\text{OH}^- + \text{H}_2$	$2\text{H}_2\text{PO}_2^- + 2\text{OH}^- \rightarrow 2\text{H}_2\text{PO}^- + \text{H}_2 + 2e^-$ $\text{H}_2\text{PO}_2^- + 2\text{OH}^- \rightarrow \text{H}_2\text{PO}^- + \text{H}_2\text{O} + 2e^-$ $\text{BH}_4^- + 4\text{OH}^- \rightarrow \text{BO}_2^- + 2\text{H}_2\text{O} + 4e^-$	4–5 or 8–10
sodium borohydride ^e	$M^{z+} + xe^- \rightarrow M^0$ $\text{B(OH)}_4^- + 3e^- \rightarrow \text{B}^0 + 4\text{OH}^-$ $2\text{H}_2\text{O} + 2e^- \rightarrow 2\text{OH}^- + \text{H}_2$	$\text{BH}_4^- + 8\text{OH}^- \rightarrow \text{BO}_2^- + 6\text{H}_2\text{O} + 8e^-$ $\text{B(OH)}_4^- \rightleftharpoons \text{BO}_2^- + 2\text{H}_2\text{O}$	12–14
aminoboranes ^{e,f}	$M^{z+} + xe^- \rightarrow M^0$ $\text{B(OH)}_4^- + 3e^- \rightarrow \text{B}^0 + 4\text{OH}^-$ $2\text{H}_2\text{O} + 2e^- \rightarrow 2\text{OH}^- + \text{H}_2$	$2\text{R}_2\text{NHBH}_3 + 8\text{OH}^- \rightarrow 2\text{R}_2\text{NH}^- + 2\text{BO}_2^- + 4\text{H}_2\text{O} + 3\text{H}_2 + 6e^-$ $\text{R}_2\text{NHBH}_3 + 7\text{OH}^- \rightarrow \text{R}_2\text{NH}^- + \text{BO}_2^- + 5\text{H}_2\text{O} + 6e^-$ $\text{B(OH)}_4^- \rightleftharpoons \text{BO}_2^- + 2\text{H}_2\text{O}$	5–10

^a M = Cu, Ni, Co, and alloy deposits unless otherwise noted.
^b M = Cu, Ag, Au.
^c Deposits are nearly pure metal.
^d Deposits contain elemental phosphorus.
^e Deposits contain elemental boron.
^f Hydrolysis of R₂NHBH₃ is the first step.

Electroless reactions must be autocatalytic. Some metals are autocatalytic, such as iron, in electroless nickel. The initial deposition site on other surfaces serves as a catalyst, usually palladium on noncatalytic metals or a palladium–tin mixture on dielectrics, which is a good hydrogenation catalyst (20,21). The catalyst is quickly covered by a monolayer of electroless metal film which as a fresh, continuously renewed clean metal surface continues to function as a dehydrogenation catalyst. Silver is a borderline material, being so weakly catalytic that only very thin films form unless the surface is repeatedly catalyzed; newly developed baths are truly autocatalytic (22). In contrast, electroless copper is relatively easy to maintain in an active state; commercial film thicknesses vary from <0.25 to 35 μm or more.

Catalysts for dielectric surfaces are more complex than the simple salts used on metals. The original catalysts were separate solutions of acidic stannous chloride [7772-99-8], used to wet the surface and deposit an adherent reducing agent, and acidic palladium chloride [7647-10-1], which was reduced to metallic palladium by the tin. This two-step catalyst system is now essentially obsolete. One-step catalysts consist of a stabilized, pre-reacted solution of the palladium and stannous chlorides. The one-step catalyst is more stable, more active, and more economical than the two-step catalyst (21,23). A separate acceleration or activation solution removes loose palladium and excess tin before the catalyzed part is placed in the electroless bath, prolonging bath life and stability.

Equipment

Plating Tanks. An electroless plating line consists of a series of lead-lined (for plastics etching) or plastic-lined tanks equipped with filters and heaters, separated by rinse tanks (24). Most metal tanks, except for passivated stainless steel used for electroless nickel, cannot be used to hold electroless plating baths because the metal initiates electroless plating onto itself. Tank linings must be stripped of metal deposits using acid at periodic intervals.

Heaters. The preferred methods for heating are a double-jacket heated tank, nonmetallic heat exchangers, quartz heaters, or Teflon-coated low watt density stainless steel heaters. Localized overheating must be avoided.

Process Control. Some hot nickel and flash electroless copper solutions are plated to the point of exhaustion and then discarded. Most baths are formulated to give bath lives of >6 turnovers of the bath constituents; some reach steady-state buildup of the by-products and can be used indefinitely. All regenerable solutions should be filtered to remove particulates that can cause deposit roughness and bath instability.

Replenishment should be done with care. Massive additions can cause decomposition. Maximum stability of electroless baths is obtained when continuous replenishment is practiced. Colorimetric analyzers are commonly used to control the addition of replenisher solutions in a set ratio based on the nickel or copper content of the bath. A number of machines are available that continuously analyze plating baths and make additions based on each separately analyzed component.

Rack Design and Loading Factors. Racks are of stainless steel, titanium, or plastic-coated steel. Tank and rack design are less important than tank loadings and rack coatings, especially when plating nonconductors. The parts may be closely racked without harm if reracking is done after electroless plating; otherwise, racking should be the same as for electrolytic plating. However, more racks may be placed in a given electroless tank than in a similar electrolytic tank. The critical points are to use air sparging to maximize metal transfer rates; filter to remove catalytic nuclei before they grow; and not to overload the plating bath. Electroless baths rapidly become overactive and decompose when too much surface area is being plated at once. Most vendors recommend a maximum loading of <245 cm² L.

Two sets of racks were originally used when plating nonconductors, reracking after the electroless bath, because all rack surfaces were plated. The racks were stripped and reused for each cycle. Racks for plating without reracking have standard copper rack splines and stainless contacts. These are coated with a special poly(vinyl chloride) plastisol that absorbs large amounts of chromic acid [11115-74-5], CrO₃, from the etchant used in plating on plastics. The chromic acid cannot be completely removed during the subsequent rinsing and neutralization steps, inhibiting deposition of electroless nickel and electroless copper on its surface.

Safety and Waste Disposal

Electroless plating operations must comply with EPA and local waste treatment guidelines. The most troublesome chemicals are electroless nickels and coppers, and the chromic acid etchant. Most solutions are considered hazardous, as they are usually highly acidic or basic, and usually contain metals (see HAZARDOUS WASTE TREATMENT). All reducing agents should be stored separately from oxidizing substances such as chromic acid or nitric acid. Proper ventilation and fume scrubbing are extremely important, as is the use of proper safety equipment. Formaldehyde reducing agent is a suspect carcinogen. All personnel must have appropriate training in case of spills or accidents.

Waste minimization (see WASTE REDUCTION) techniques must always be implemented as part of waste disposal and treatment. Countercurrent rinsing reduces water flow and facilitates waste treatment. An example is the control of plastic etching baths. Stagnant rinses are used after etching to further reduce chromic acid losses. Most installations use an electrolytic regeneration cell for chromic acid etchants. The cell consists of a double chamber containing a porous pot holding a lead anode where Cr³⁺ is continuously oxidized to Cr⁺ and recycled to the etch tank. The Cr³⁺ content must be kept at a fairly low concentration for good etching. The chromic acid etchants do not need to be dumped; thus only small amounts from drag-out losses must be waste-treated. These are often recovered by evaporation.

The electroless baths, especially copper and nickel, may be difficult to treat by conventional precipitation because of the presence of strong chelating agents (qv). Ion-exchange and accelerated plateout systems are among the best techniques, although ozonolysis and other oxidative treatments also destroy chelating agents. The metal is tightly chemically bound by organic complexers or chelating agents. Specific techniques for breaking the complex and precipitating the metal include treatment using lime, ferrous sulfate, sodium borohydride [16940-62-2], and organic sulfides. Suppliers of the proprietary baths can make specific recommendations, depending largely on the type of chelate used.

Electroless silver solutions must be specially neutralized before treatment because a dried silver–amine complex can explode; newer technology has eliminated this problem.

Adequate ventilation is necessary for all process lines to ensure worker safety. Electroless copper baths must have good ventilation to remove toxic formaldehyde vapors and caustic mist generated by the hydrogen evolution reactions and air sparging. Electroless nickels need adequate ventilation to remove nickel and ammonia vapors. Some states and municipalities require the removal of ammonia from wastewaters. A discussion of printed circuit board environmental issues and some sludge reduction techniques is available (25).

Plating on Metals

Electroless Nickel. The first large-scale electroless process was the Kanigen electroless nickel process from General American Transportation Corp. This hot nickel process uses a hypophosphite reducing agent. Properties of electroless nickel deposits vary greatly depending on the reducing agent: hydrazine [302-01-2] gives a practically pure nickel [7440-02-0]; the organoboron reducing agents give very hard nickel–boron alloys; and the most widely used hypophosphite deposit a range (from 1–15 wt % P) of nickel–phosphorus alloys having unique properties. Acidic (pH 4.0–5.5) baths are preferred, but alkaline (pH 8–10) baths are also used. Operating temperatures are 70–95°C. Typical formulations use sodium hypophosphite [7681-53-0] as the reducing agent, nickel sulfate [7786-81-4] or nickel chloride [7718-54-9], and ammonium or organic salts as buffers. Mild complexing agents such as citrate or glycolate, and trace quantities of proprietary stabilizers including heavy metals or sulfur compounds, are used. A large number of commercial baths are available (7,26,27).

Electroless nickel–boron baths use sodium borohydride or dimethylamine borane [74-94-2] in place of sodium hypophosphite (see BORON COMPOUNDS). The nickel–boron alloy is brittle, highly stressed, and much more expensive than nickel–phosphorus alloys. Nickel–boron is mainly used to replace gold in printed circuit board plating.

The engineering properties of electroless nickel have been summarized (28). The Ni–P alloy has good corrosion resistance, lubricity, and especially high hardness. This alloy can be heat-treated to a hardness equivalent to electrolytic hard chromium [7440-47-3] (Table 2), and the lubricity is also comparable. The wear characteristics are extremely good, especially with composites of electroless nickel and silicon carbide or fluorochloropolymers. Thus the main applications for electroless nickel are in replacement of hard chromium (29,30).

Table 2. Taber Abraser Resistances of Engineering Coatings^a

Coating	Heat treatment ^b	TWI ^c
Watts nickel	none	25
electroless nickel		
9% phosphorus	none	17
	300°C	10
	500°C	6
	650°C	4
5% phosphorus	none	9
	400°C	3
hard chromium	none	3

^a Courtesy of MacDermid Inc., Waterbury, Conn.

^b For time period of 1 h.

^c TWI = Taber Wear Index. CS-10 abramer wheels, 100 gram load, determined as average weight loss per 1000 cycles for total test of 6000 cycles.

The advantages of electroless nickel over hard chromium include safety of use, ease of waste treatment, plating rates of as much as 40 μm/h, low porosity films, and the ability to uniformly coat any geometric shape without burning or using special anodes. Increased chemical safety is another

advantage, as hexavalent chromium, which is used for hard chrome plating, is being very tightly regulated because of its carcinogenicity. Chromium metal, however, is not toxic. The uniformity of coating thickness of electroless nickel can also minimize expensive machining after plating, and can be used to salvage parts that have been overmachined. Electroless nickel has superior corrosion and erosion resistance as compared to electrolytic nickel. It is used extensively on molds, pistons, pump parts, oil field equipment, dies, compressors, tanks (qv), and piping (see PIPING SYSTEMS).

The market size for electroless nickel solutions is not known with certainty. An estimate for 1990 is 1200 t of nickel, of which 85° may be for plating on metals, and 15° for plating on nonconductors. This estimate is based on hypophosphite production, because virtually all hypophosphite is used for electroless nickels. The corresponding market value of the materials used for plating on metals was \$35–40 million. Plating of nonconductors is at least 10 times as great in terms of area plated owing to the much lower thicknesses used as compared to metal plating.

Electroless nickel can be used to plate aluminum [7429-90-5], Al. The adhesion is often poor unless the aluminum is etched to remove oxides. The best method is to use an intermediate zincate deposit. Adhesion can vary widely depending on the aluminum alloy used.

Electroless nickel or nickel–lead alloys can improve the solderability and braisability of aluminum even when a continuous film is not present. Electroless nickel systems based on dimethylamineborane reducing agents are used to coat aluminum contacts and semiconductors (qv) in the electronics industry. Newer uses include corrosion-resistant electroless nickel topcoatings on electroless copper plating for radio frequency interference electromagnetic interference (RFI EMI) shielding, and as an undercoating for electroless gold deposits. Many types of composite electroless nickel coatings are now available. The included phase ranges from diamond [7782-40-3] to Teflon to aluminum oxide [1344-28-1], Al₂O₃, and aluminum nitride [24304-00-5], AlN.

Substrate Preparation. All parts must be cleaned thoroughly before plating. Any traces of dirt or oxide either prevent deposition or lead to adhesion loss. Nickel can be spontaneously deposited from hot electroless solutions on most common metals, including mild steel, beryllium–copper, aluminum, stainless steel, and titanium. Generally, special etchants are used for difficult materials such as titanium and some stainless steels. Lead and a few other materials cannot be plated directly because these poison the electroless reaction. However, a thin electrolytic layer may be used to mask such surfaces to make them platable. Where initiation is slow or difficult, such as for copper substrates, plating can be started by briefly contacting the inactive part with one that is actively plating, by giving the surface a brief cathodic pulse of current, by contact with a dissimilar metal, such as iron, or by prior immersion in a dilute solution of a precious metal such as platinum or palladium.

Parts are allowed to plate until the desired thickness is reached. This point can be monitored by knowledge of the bath operating rate, or by direct thickness or weight measurement of standard test parts.

Other Metal Processes. The only other types of electroless processes commonly used on metals are electroless gold and palladium. These are restricted to specialized uses because of extremely high cost. The properties of these coatings are not significantly different from electroplated coatings. Organoboron, hypophosphite, and other reducing agents are used in proprietary baths. Electroless coatings are used where noncontinuous metal base materials make electrolytic coating impossible. Electroless gold baths in particular have suffered from instability and short bath life, but newly developed solutions have overcome this problem (31). Typical applications include solder mask-coated printed circuit boards, insertion tabs, and surface mount devices. New electroless silver and lead processes are available for specialized uses.

Plating on Nonconductors

Plating on nonconductors comprises two technologically very different categories: plating of plastics and printed circuit production.

Plating of Plastics. Plastics can be coated with metals for decorative and or functional purposes. Plating on plastics (POP) was first used commercially about 1905 to plate celluloid pen parts. An electroless silver solution was applied to the surface after stannous chloride sensitization. The silver film was used as a base for further electroplating. This process was based on an 1858 patent for production of silver deposits on mirrors using electrolytically plated metal (32). Further use of POP was limited mostly to encapsulation processes such as electroless silver button plating until the early 1960s when first-generation electroless coppers were used. By the mid-1960s, electroless nickel formulations were developed using pH 9–10 that could be used for plating of plastics at room temperature (9).

Most molded plastics have a very smooth, hydrophobic surface that must be modified. Chemical etchants are used to oxidize and roughen the surface. The resultant hydrophilic surface promotes good metal-to-plastic adhesion. The etchant is usually a solution of chromic acid and sulfuric acid; pure chromic acid can also be used.

As of 1992, the first specialty platable plastic, acrylonitrile–butadiene–styrene (ABS) terpolymer (see ACRYLONITRILE POLYMERS, ABS RESINS), is used in over 90° of POP applications. Other platable plastics include poly(phenylene ether) (see POLYETHERS), nylon (see POLYAMIDES), polysulfone (see POLYMERS CONTAINING SULFUR), polypropylene, polycarbonate, phenolics (see PHENOLIC RESINS), polycarbonate–ABS alloys, polyesters (qv), foamed polystyrene (see STYRENE PLASTICS), and other foamed plastics (qv).

Electroless films have two functions: they provide an electrically conductive layer that allows further coating by electroplating; and they provide a secure bond between the plastic and the electroplated layer. Plated plastics, however, have several disadvantages. Plating normally lowers impact strength. The coefficient of thermal expansion is much higher for plastics than for metals, so stress buildup and adhesion loss can occur on severe thermal cycling. Blistering can occur during corrosion. The relatively low heat distortion temperature of most plated plastics can also limit applications. Molding cycles are relatively long (33).

One of the most important advantages of plated plastics is that weight savings can be as much as 60° as compared to an equivalent all-metal part. The molded plastic parts need no buffing or other finishing step before plating. Plated plastics have improved tensile strength, elasticity, flexural strength, a reduced total coefficient of thermal expansion, and improved abrasion and weathering resistance. Whereas a metal part may completely corrode away and fail in service, plastic gives extended life because only the surface of a plated plastic can be corroded. This assumes that the part design was adequate, the proper plating-grade resin was selected, and that the molding conditions were correct. All of the principal suppliers of plating-grade resins offer detailed help on these problems.

Applications and Economic Aspects. There are various reasons for coating plastics with metals. In the automotive industry, the light weight of plastics is combined with the consumer appeal of metal. The largest single application for electroless plating is for automotive parts such as grilles and headlamp bezels. This market segment has dropped in size because cars are smaller, and painted parts have replaced painted finishes. The electroless coating serves only as the base for electrolytic final metal coatings. Other large market segments include hardware, appliances, marine applications, plumbing fixtures, knobs, closures, and novelties; RFI EMI-shielding applications are a newer market opportunity (34). Plated ABS has about 80° of the market. Most of the remainder is ABS–polycarbonate alloys (especially for computer housings) and modified poly-(phenylene oxide).

As of 1990, plastics are plated in fewer than 50 U.S. plants. The annual electroless chemical costs range from \$7–10 million for POP, and \$5–7 million for RFI shielding. Individual plants process from 4.6 × 10³ to >1.3 × 10⁵ m² yr of plastic surface area. Automotive items make up over 50° of the market on a plated area basis; the remainder is hardware, plumbing, electronics shielding, and decorative items.

Process Details. Automated processing lines allow low cost production of plated parts. Rejects, which are typically less than 5° of production, can be stripped and replated, or reground and remolded. The cost of chemicals is typically \$1.07–2.15 m².

The process consists of pre-etching, etching, etch neutralization, catalyst application, catalyst activation, and plating. Most commercial applications, except RFI EMI shielding, use the initial copper or nickel deposit as a base for subsequent electrolytic plating of electrolytic copper, nickel, or chromium. The exact types and thicknesses of metal used are determined by part usage, eg, automotive exterior, decorative, plumbing, and others (24).

Chemistry. Successful electroless plating depends on the optimized interaction of five separate complex chemical solutions (1) to clean, roughen, and catalyze the surface before plating. These steps are critical for formation of an adherent continuous electroless coating, and for optimum durability after electrolytic plating.

Etching is necessary to give good metal-to-plastic adhesion. Solutions of strongly oxidizing chromic acid–sulfuric acid–water, or chromic acid–water are operated near the point of mutual saturation. The etchant both physically roughens the surface and chemically modifies it to give a very hydrophilic surface. Metal adhesion occurs owing to the combined effects of chemical bonding and mechanical locking to the roughened surface. In the etching of

ABS plastic, a polymer consisting of polybutadiene spheroids is dispersed in a continuous phase of poly(styrene–acrylonitrile). The chromic acid attacks the polybutadiene at a much higher rate than the continuous phase. This gives an excellent microroughened surface with superior metal-to-plastic bond strength. A typical recommended formulation consists of 20 vol % sulfuric acid, 420 g L chromic acid, and 0.1–1.0% of a fluorocarbon wetting agent. The plastic is treated with this formulation for 6–10 min at 60–65°C.

Neutralizing removes the large amount of hexavalent chromium from the surface of the part. Hexavalent chromium shortens the life of the catalyst, and trace amounts completely inhibit electroless nickel deposition. The neutralizer is usually a mildly acidic or basic reducing agent, but other types of neutralizers are available, especially for substrates that are difficult to plate. The neutralizer may also contain surfactants (qv) or other compounds that increase catalyst absorption; absorption promoters are often needed for non-ABS plastics.

Catalysis is done by an acidic solution of the stabilized reaction product of stannous chloride and palladium chloride. Catalyst absorption is typically 1–5 µg Pd per square centimeter. Other precious metals can be used, but they are not as cost-effective. The exact chemical identity of this catalyst has been a matter of considerable scientific interest (19–21,23). It seems to be a stabilized colloid, co-deposited on the plastic with excess tin. The industry trends have been to use higher activity catalysts at lower concentrations and higher temperatures. Typical usage is 40–150 ppm of palladium at 60°C maximum, and a 30–60-fold or more excess of stannous chloride. Catalyst variations occasionally used include alkaline and non-noble metal catalysts.

Acceleration modifies the surface layer of palladium nuclei, and stannous and stannic hydrous oxides and oxychlorides. Any acid or alkaline solution in which excess tin is appreciably soluble and catalytic palladium nuclei become exposed may be used. The activation or acceleration step is needed to remove excess tin from the catalyzed surface, which would inhibit electroless plating. This step also exposes the active palladium sites and removes loose palladium that can destabilize the bath. Accelerators can be any acidic or alkaline solution that solubilizes excess tin.

Electroless metal is provided by either copper or nickel. A number of reports have indicated that copper may be better for corrosion resistance (33,35). A nickel–copper–phosphorus alloy of low copper content is also available (27). Although the electroless copper plating systems are similar to those used in printed circuit processing, electroless nickel processes differ greatly from those used in metal plating and are not interchangeable. Room temperature nickel solutions are mildly alkaline, and give a low and uniform deposition rate. Automotive and other heavy-duty applications use copper, whereas decorative and interior articles are often plated with nickel. RFI EMI shielding applications use nickel–phosphorus for low shielding requirements, or a copper layer coated with nickel or other metal for high shielding (36).

Room temperature electroless nickel systems are easily controlled and replenished with indefinite bath lives. Electroless copper solutions can approach these bath lives when carefully controlled. As only thin-metal layers are deposited, the drag-out losses can be comparable to the actual usage of metal in plating. Drag-out, and thus the amount of metal lost, is a constant per surface area; the thinner the plated deposit, the greater the ratio of metal drag-out loss to metal usage. This drag-out helps bath stability by keeping the reaction products from building up.

A plating time of 5–20 min gives a 0.15–0.5 µm film that allows electroplating of the part as though it were totally metallic. Plated parts, except those for RFI EMI shielding, are electroplated with multiple coats of copper, nickel, and chromium. Precautions must be taken to control the initial current during electroplating, or the thin electroless plated film may overheat and burn off. A nickel or copper strike is usually applied first, followed by any desired decorative or functional coating. Special stop-off paints are commonly applied to the backs of large auto parts, such as grilles, to confine metal plating to appearance surfaces only. Types and thicknesses of coatings depend on the end use of the part. They range from 22.25 µm total Cu + Ni + Cr for mild service conditions to 45.25 µm for severe conditions. Recommended minimum electroplate thickness for a bright chromium finish are bright acid copper, 15 µm; semibright nickel, 20–30 µm; or bright nickel, 7–15 µm; and chromium 0.25 µm (37). Details of electrolytic plating can be found elsewhere (24,27,38).

Specifications and Test Methods. The American Society of Electroplated Plastics publishes a comprehensive book covering most technical aspects of testing and control (24). A number of ASTM standards have been issued that cover thickness, adhesion, and thermal-cycling resistance of the total plated film. Some specifically for plated plastics include:

ASTM test number	Description
B117	salt spray test
B287	acetic acid salt spray test
B380	corrodkote test
B368	CASS test
B532	recommended practice for evaluation of appearance
B533	peel strength measurement of plated plastics
B553	thermal cycle test
B554	recommended practice for measurement of metallic coating thickness on nonmetallic surfaces
B571	adhesion of metallic coatings
B604	decorative electroplated coatings for plastics
D2247	humidity test

Additional ASTM tests pertain to covering over electroless coatings, such as hardness of electroless nickel.

The physical properties of plated plastics may be quite different from those of molded parts; tensile strength, elastic modulus, flexural strength, and heat-distortion temperature are usually higher. Corrosion resistance is generally better than that of a steel or zinc part. The thermal coefficient of expansion is lower, however, and this can cause stress cracking if not properly compensated for in design. Ductility and impact strength are lower (33). Underwriters' Laboratory also sets standards for plated plastics, including those used for RFI EMI shielding, under UL 746C. In addition, the principal auto producers have their own sets of standards.

Printed Circuit Boards. Electroless copper metallization is an indispensable part of the modern electronics revolution. Printed circuit board production uses electroless copper to coat nonconductive plastic surfaces exposed when drilling through-holes for internal electrical paths, and to define the circuit patterns on the board surfaces. Printed circuit production is much more complicated than POP, involving up to 80 or more process and inspection steps. Automated electroless plating lines are often used, but separated from the electrolytic processing cycle. Thus the boards are processed and plated in the semibulk mode, where large numbers of boards are spaced closely together on special racks. The electroless plated panels are reracked for electrolytic plating after the intermediate processing steps. Printed circuit production also includes electrolytic plating of copper, tin, tin–lead, nickel, and gold.

Numerous variations exist in the electroless plating solutions, processes, and techniques employed both in laboratory and commercial form, to create a great variety of products (39). All produce a layer of highly conductive copper in specified areas. Modern electroless copper films have a ductility and conductivity identical to that of electrolytic copper (40). The three basic classes of copper baths are

Copper bath	Thickness desired, µm	Processing type
flash	0.5–1	subtractive
fast	1–2.5	semiadditive
heavy build	25–40	additive

The processing type refers to how the copper circuits are fabricated. These classes differ in stability, ease of regeneration and control, operating

temperature, cost, and deposition rate. Subtractive processes dominate commerce; additive baths are used by larger, specialized processors; and the semiadditive processes show little use.

Most printed circuit board (PCB) production uses the subtractive process (41). In the simplest version, a thin copper foil is laminated to a nonconductor, holes are fabricated, and the unwanted copper etched off. These single-sided boards do not require plating. Known as print-and-etch, this version is used for the most simple printed circuit boards.

The first commercial boards using electroless plating for connection between the sides were made by Motorola in 1953 (41). An unclad plastic board was coated with electroless silver, a reverse resist applied, and copper electroplated to the required circuit thickness. The resist and excess silver were removed. This semi-additive process is still used, with copper replacing the silver.

The need to install more components on one board led to the development of double-sided boards having copper laminated to both sides of a nonconducting support. Techniques were developed to apply electroless copper baths simultaneously to copper foil and bare plastic holes on a composite board. The copper thickness in the holes is then increased by electrolytic copper plating either before circuit definition using a mask (panel plating) or after (pattern plating), followed by final copper etching. This process, known as the subtractive process, is used for most production as of this writing. Multilayer boards having increased component-packing densities are a further refinement. The boards are fabricated from alternating layers of plastic insulator and copper foil. The inner copper foil layers are etched to define internal circuit paths before lamination. The alternating copper-plastic layers are assembled into a stack, laminated, and drilled. Layers of up to 20 or more copper foils interleaved with insulators are being commercially prepared.

There has been a continual increase in size and complexity of PCBs with a concurrent reduction in conductor and hole dimensions. Conductors can be less than 250 μm wide; some boards have conductors less than 75 μm wide. Multilayer boards greater than 2.5 mm thick having hole sizes less than 250 μm are being produced. This trend may, however, eventually cause the demise of the subtractive process. It is difficult to etch such fine lines using 35- μm copper foils, though foils as thin as 5 μm are now available. It is also difficult to electroplate holes having high aspect ratio. These factors may shift production to the semiadditive or fully additive processes.

The dielectric separating layer in PCBs may be plastic or another nonconductor, eg, ceramic or ceramic-coated steel (see CERAMICS). The plastic must be dimensionally, thermally, and chemically stable under operating conditions, have good flex strength, be resistant to humidity, have low thermal coefficient of expansion, be fire retardant, and have a high dielectric strength and a low dielectric constant. No one material is ideal, but the largest volume consists of various types of epoxy resins (qv) impregnated onto woven glass fiber and partially cured. These are then cut to size, stacked in layers alternating with copper-foil layers, and heated under pressure for full cure and bonding. Other substrates include phenolic-paper, epoxy-paper, polyester-glass, and polyimide-glass for standard processing. High frequency applications may require polytetrafluoroethylene, polystyrene, or cross-linked polyethylene cores.

Flexible PCBs are produced in large quantities, usually consisting of polyester, polyimide, and polyimide-Teflon films. Both foil-laminated and uncoated versions are available. A modern complex multilayer board after final fabrication is shown in Figure 2.

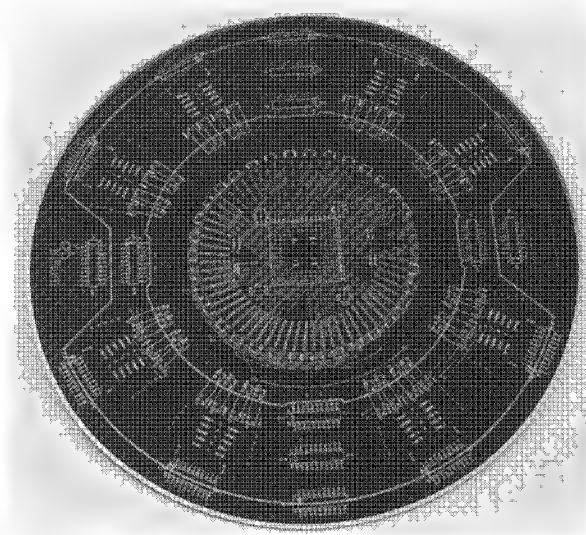


Fig. 2. Multilayer printed circuit board composite. Construction is multiple layers of epoxy-glass and foil copper. Foil copper outermost layer and drilled through-holes are sequentially plated with electroless copper, electrolytic copper, electroless nickel, and electroless gold.

Courtesy of Applied Electroless Concepts, Inc.

Newer resins include polysulfone, polyethersulfone, polyetherimide, and polyetherketone. Some of these newer materials are high temperature thermoplastic, not thermoset, resins. They are being promoted for the design of injection-molded printed circuit boards in three-dimensional shapes for functional applications as an alternative to standard flat printed circuit boards. Only semiadditive or fully additive processing can be used with these devices.

Economic Aspects. Printed circuit board production is a multibillion dollar business in the United States alone. An estimated 800 shops are involved in production, ranging from small companies specializing in one aspect of production, to huge organizations containing both research and production facilities capable of producing any type of board. U.S. production of printed circuit boards reached \$5.4 billion in 1990, an increase from \$3.5 billion in 1983. Increase in worldwide production was much greater because of massive shifts of production to overseas facilities and closure of over half the U.S. shops. Electroless plating is only a small part of this cost (42).

The cost of the chemicals used in electroless copper plating is very low, rarely exceeding \$2.78 m^2 , except for fully additive processes. The principal costs of printed circuit board production arise mainly from handling steps and other operations.

In 1990 the majority of U.S. PCB production resulted from subtractive or print-and-etch processing; additive processes were less than 6% of the total; multilayer boards accounted for 55.8%. The ratio of rigid to flexible surface areas plated is about 15:1. High performance plastics including polyimide, Teflon, and modified epoxy comprised 6% of the market (\$324 million); flexible circuits were 6.6% (\$360 million) (42).

Multilayer boards, which use multiple interior laminates of plastic and copper, now comprise over half of the value of production, though much less on a surface area basis. Surface mount technologies demand extreme flatness and reproducibility from surfaces. Greater packing density has led to commercial production of finer lines and holes, often less than 50 μm and 500 μm , respectively. Electroless gold over electroless nickel-phosphorus, or electroless nickel-boron alone, is often used as a topcoating for wire bonding or improved solderability.

Chemistry. Many of the chemicals and process steps used for PCB processing are similar to those used in the POP industry; however, electroless copper is used exclusively (39,41). The surface must be clean for uniform coverage, surface roughening, and copper deposition. Copper-clad boards are given an acid or alkaline cleaning step. The copper layers are then micro-etched with acidic persulfate or hydrogen peroxide-sulfuric acid solutions to promote adhesion of the electroless copper. Multilayer boards receive a precleaning step to eliminate resin smear on the internal copper connections. The precleaning step may use plasma etching (see PLASMA TECHNOLOGY), concentrated sulfuric acid, alkaline potassium permanganate, or concentrated chromic acid. The mixed chromic acid-sulfuric acid etchants used for POP cannot be used, as these rapidly dissolve copper foils.

The cleaned and microetched boards are catalyzed and activated in the same manner as for POP. The same coppers are used as for POP, but these are usually formulated to give greater deposition rates and thicknesses. Copper is typically deposited in 0.25–35 μm thickness, depending on the process. Fully additive processing needs thicknesses of 25–35 μm; subtractive processing needs thicknesses of 0.25–2.5 μm.

Application Methods. The three PCB processing techniques may be summarized as:

Subtractive process	Semiadditive process	Additive process
copper-clad board	thin copper-clad board	unclad board
form holes	form holes	adhesive coat
catalyze	catalyze	form holes
accelerate	accelerate	etch surface
electroless copper	electroless copper	catalyze
positive resist coat	negative resist coat	negative resist coat
electrolytic copper	electrolytic copper	accelerate
dissimilar metal electroplate	dissimilar metal electroplate	electroless copper
strip resist	strip resist	
etch copper	etch copper	

Additive and semiadditive processing can give material savings and higher circuit densities, whereas subtractive processing is technologically easier.

Additional steps used for multilayer production include application of black or red oxide to foil; application of photoresist or screened etch resist; imaging and development of photoresist; copper etching; removal of resist; assembling of layers, checking registration and lamination; drilling; and removal of drill smear on copper inner layers. Rinses and additional cleaning, neutralizing, light etching, and developing steps are required (39).

Subtractive Method. Technologically, the subtractive process is the least demanding. Copper foil (31 mg cm²) is laminated to both sides of an insulator such as an epoxy–glass composite (see COMPOSITE MATERIALS, POLYMER MATRIX). Through-holes are punched or drilled. The board is cleaned, catalyzed, and placed in a flash electroless copper bath. A thin conductive layer of copper deposits on the entire board, including the nonconductive holes, to a thickness of 0.25–2.5 μm. The copper thickness is increased to 25–35 μm by electroplating. After suitable masking, the unwanted copper foil is etched away, leaving the defined circuit paths. The main disadvantage is the large amount (as much as 50–80% of the original copper) of copper foil wasted. Uniform thickness, conductivity, and adhesion of the copper are critical.

Semiadditive Method. The semiadditive method was developed to reduce copper waste. Thin 5.0 μm (4.5 mg cm²) copper foil laminates are used, or the whole surface may be plated with a thin layer of electroless copper. Hole forming, catalysis, and electroless copper plating are done as for subtractive circuitry. A strippable reverse–resist coating is then applied. Copper is electroplated to 35 μm or more, followed by tin or tin–lead plating to serve as an etch resist. The resist is removed, and the whole board is etched. The original thin copper layer is quickly removed to leave the desired circuit. This method wastes less than 10% of the copper.

Fully Additive Method. No electrolytic plating step is used in the fully additive process. The copper circuit is formed directly on the board without a continuous copper film. Heavy-build electroless coppers are used to increase the final thickness of the entire circuit. This process is much more difficult to control than the others. Additive processing is becoming increasingly important in high aspect ratio, very small diameter through-holes that cannot be easily electrolytically plated.

Special unclad laminates are used that have been coated with an easily etchable adhesion layer, typically an epoxy paint containing a polybutadiene dispersion. A solvent solution may be used to swell the surface and facilitate etching. The treated laminate is etched in a chromic acid–sulfuric acid solution of the type used for plating on plastics. The boards are processed as for the semiadditive method, but removed and dried after the catalyst is applied. A resist is applied wherever a circuit is not wanted. The boards are accelerated and placed in a special electroless copper solution for 16–30 h. A full 25–50 μm thickness of copper is deposited by electroless plating only. Electrolytic copper plating cannot be used because there is no continuous conductive film linking the discrete circuit paths. No copper etching step is needed and no copper is wasted.

Electroless Copper. Electroless copper, introduced in the mid-1950s (41,43), is available commercially in great variety. Formaldehyde is usually the reducing agent, copper sulfate (occasionally copper nitrate or copper chloride) is the metal salt, and sodium hydroxide is used to control pH. The complexers may be tartrate, ethylenediaminetetraacetic acid (EDTA), tetrakis(2-hydroxypropyl)ethylenediamine, nitrilotriacetic acid (NTA), or some other strong chelate. Numerous proprietary stabilizers, eg, sulfur compounds, nitrogen heterocycles, and cyanides (qv) are used (2,44). These formulated baths differ in deposition rate, ease of waste treatment, stability, bath life, copper color and ductility, operating temperature, and component concentration. Most have been developed for specific processes; all deposit nearly pure copper metal.

Sodium borohydride or dimethylamine borane have found limited use as reducing agents because of expense. In addition, bath stability, plating rate, and deposit properties are inferior to those of formaldehyde-reduced baths. The deposit is a copper–boron alloy. Copper–hypophosphite baths have been investigated, but these are poorly autocatalytic, and deposit only very thin coatings.

Printed Circuit Etchants. Two types of etchants are used in printed circuit board processing for entirely different purposes. Etchants used for etchback or desmear attack only organic materials, whereas copper etchants dissolve only copper. Etchback refers to massive removal of the organic board material perpendicular to the drilled hole, to expose annular copper rings from the inner layers of multilayer boards to ensure good conductivity after plating. Desmear is a less aggressive type to roughen drilled hole surfaces and remove organic residues from the drill-exposed copper inner layers. Older chromic acid and sulfuric acid etchants have been replaced by regenerable permanganate systems and by plasma etching. These etchants directly contribute to electroless plating through micro-roughening the surface and removing loosely bonded debris, thereby improving adhesion and performance.

Copper etchants do not directly influence the electroless plating process, but are used merely to remove unwanted copper, and should not affect the deposit properties. The costs of waste treatment and disposal have led to disuse of throw-away systems such as chromic–sulfuric acid, ferric chloride, and ammonium persulfate. Newer types of regenerable etchants include cupric chloride, stabilized peroxide, and proprietary ammoniacal etchant baths.

Emerging Printed Circuit Technologies. Much research has been devoted to developing safer, cheaper, and more reliable alternatives to formaldehyde-reduced electroless coppers, and many processes are in the advanced stages of development (45). These include hypophosphite-reduced electroless coppers; substitution of conductive graphite colloids for electroless coppers; direct electroplating over catalyzed or catalyzed and sulfide treated substrates; and non-noble metal catalysts. Three-dimensional printed circuit boards, usually on engineering quality plastics, are becoming more common (see ENGINEERING PLASTICS). High performance boards of novel types continue to be developed with controlled impedance layers; ultrathin and narrow copper conductors; ceramic substrates; high temperature plastic laminates; and Invar or Kovar inner layers in place of copper or dielectric.

Other Processes

Electroless Silver. Silver, the first metal used for metallizing nonconductors, is rarely employed because of high cost, low stability, and inability to plate a thick deposit autocatalytically. Formulations typically consist of ammoniacal silver nitrate solutions and a reducing agent of sugar, formaldehyde, hydrazine, or dimethylamineborane (3,13,46). These baths are generally unstable, and are used until exhausted. The silver–amine complex must not be allowed to dry without neutralization because it is shock-sensitive and detonates. The two solutions are mixed just before use, and the parts immediately placed in the solution; plate adhesion is often poor. A catalyst is normally not required, though a sensitizing solution of acidic stannous chloride gives better coverage and adhesion. Some formulations are so active that they are mixed in a spray head, sprayed on a part, and discarded. The deposit consists of nearly pure silver of high conductivity, and serves as a base for further electrolytic plating steps.

Most electroless silver applications are for silvering glass or metallizing record masters. Mirror production is the principal usage for electroless silver. The glass support is cleaned, catalyzed using a two-step catalyst, and coated on one side with an opaque silver film (46). Silver-plated nylon cloth is used as a bacteriostatic wound dressing. A tiny current applied to the cloth causes slow silver dissolution. The silver acts as a bactericide (47).

Silver is also used in the jewelry and button industry. Moreover, a newer type of electroless silver has been developed by Applied Electroless Concepts that is fully stable, nonexplosive, and has none of the other traditional disadvantages (22).

Architectural Glass. The other important commercial glass-plating application is for production of architectural reflective glasses. Translucent metal films are used for decoration and for reduction of environmental heat gain. Electroless plating is used by one producer for this type of product (48).

These glass sheets are available in a wide range of optical densities and colors. The glass sheets are cleaned, coated with a two-step catalyst, and sprayed with a highly active electroless bath that deposits a very thin translucent colored film. The deposits are not thick enough to be weather- or wear-resistant, so only the inner side of the glass is coated. Electroless gold, silver, nickel, and copper are used singly or in combination to give the desired effects.

Ceramic Plating. Ceramic plating is similar to glass plating in that a two-step catalyst is used. The surface is often mechanically roughened or chemically etched to improve the metal adhesion. Ceramic resistors are usually coated with electroless nickel, although ceramic circuit boards use electroless copper. Electroless gold is beginning to have an impact, with the coatings plated over electroless or electrolytic nickel.

Composite Plating. An electroless nickel matrix can be used to securely bond diamonds to cutting tools, and electroless nickel–diamond composites are also used (see TOOL MATERIALS). The NYE-CARB process gives a silicon carbide–electroless nickel composite that has extremely high abrasion resistance (49). Electroless nickel–Teflon composites are being promoted as low friction materials.

Electroless Gold and Palladium. Electroless gold is commercially used on printed circuit boards, flex circuits, ceramics, three-dimensional circuits, metals, and highly specialized substrates. The newest type is extremely stable and cost-effective; older baths were highly unstable. A different reducing agent and associated stabilizer led to a bath that is autocatalytic and builds high thicknesses (Fig. 2) (31). Electroless palladium is also used in specialized applications (27). These baths are used mainly in specialty applications in the electronics industry where the need for pore-free or discontinuous coatings justified the high price. Both coatings are normally applied over electroless nickel, Kovar, or other materials, but cannot be used directly on copper.

Miscellaneous. The patent literature contains many unusual baths and applications, most of which have never been commercialized. The use of electroless solutions in dentistry for improving the adhesion of fillings has been described (50). Many cobalt–phosphorus and nickel–cobalt–phosphorus alloys have good magnetic properties, and have been proposed for use in computer memories and magnetic recording tape (see COMPUTER TECHNOLOGY; INFORMATION STORAGE MATERIALS) (2,51). Objects such as leaves, insects, and sea shells can be coated by encapsulation (13). Electroless platinum can be plated on ion-exchange membranes for fuel-cell applications (see FUEL CELLS). Polymer microcapsules have been electroless plated to give products of controlled permeability. Composites of electroless copper coatings have been applied to nitrocellulose propellant compositions to allow ignition by low voltage direct current. Electroless nickel and copper have been used to uniformly coat glass microspheres for tests of laser-induced hydrogen fusion (51,52) (see FUSION ENERGY).

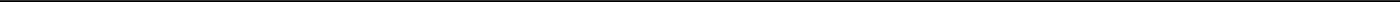
BIBLIOGRAPHY

"Electroless Plating" in *ECT* 3rd ed., Vol. 8, pp. 738–750, by G. A. Krulik, Borg-Warner Corp.

1. G. A. Krulik, *J. Chem Ed.* **55**, 361 (1978).
2. J. MacDermott, *Plating of Plastics with Metals, Chemical Technology Review No. 138*, Noyes Data Corp., Park Ridge, N.J., 1979; J. I. Duffy, ed., *Electroless and Other Nonelectrolytic Plating Techniques-Recent Developments, Chemical Technology Review No. 171*, 1980.
3. W. Goldie, *Metallic Coating of Plastics*, 2 Vols., Electrochemical Publications Ltd., Middlesex, UK, 1968.
4. F. Pearlstein and H. E. Bellis, *Electroless Plating of Metals*, American Electroplaters Society, Winter Park, Fla., 1972.
5. W. Mindt, *J. Electrochem. Soc.* **117**, 615 (1970); **118**, 93 (1971).
6. C. Singer and co-workers, eds., "The Industrial Revolution," in *History of Technology*, Vol. 4, Oxford University Press, New York, 1958.
7. *Am. Soc. Test. Mater. Spec. Tech. Publ.* 265, *Symposium on Electroless Plating*, American Society of Testing Materials, Philadelphia, Pa., 1959.
8. **U.S. Pat. 2,532,283** (Dec. 5, 1950), A. Brenner and G. Riddell (to U.S. government); A. Brenner, *Plating Surf. Finish.* **71**, 24 (1984).
9. **U.S. Pat. 4,061,802** (Dec. 6, 1977), F. E. Costello.
10. E. B. Saubestre, *Plating* **59**, 563 (1972).
11. *Electroless Nickel News* **8**(1), 2 (Spring, 1991).
12. K. M. Gorbunova and A. A. Nikiiforova, *Physiochemical Principles of (Chemical) Nickel Plating*, Academy of Sciences, Moscow, 1960 (Engl. trans. available from U.S. Government Office of Technical Services, Washington, D.C., 1963).
13. B. Y. Kaznachei, *Itoji Nanki Elektrokhim.* **196** (1965); *Chem. Abstr.* **68**, 89414n (1968).
14. P. Cavallotti and G. Salvago, *Electrochim. Metal.* **3**, 239 (1968).
15. G. G. Gawrilov, *Chemical (Electroless) Nickel-Plating*, Portcullis Press, Redhill, UK, 1979.
16. J. E. A. M. Van Den Meerakker, *J. Appl. Electrochem.* **11**, 387 (1981); **11**, 395 (1981).
17. M. Paunovic, *Plating Surf. Finish.* **55**, 1161 (1968).
18. C. I. Courdeuvelis and G. Sutcliffe, *Plating Surf. Finish.* **67**, 71 (1980).
19. G. A. Krulik, *Platinum Met. Rev.* **26**, 58 (1982).
20. E. Matijevic, A. M. Poskanzer, and P. Zuman, *Plating Surf. Finish.* **62**, 958 (1975).
21. G. A. Krulik, *J. Catal.* **65**, 95 (1980).
22. Pat. Appl. (1993), N. Mandich, G. Krulik, and R. Singh.
23. **U.S. Pat. 3,011,920** (Dec. 5, 1961), C. R. Shipley (to Shipley Corp.).
24. American Society of Electroplated Plastics, *Standards and Guidelines for Electroplated Plastics*, 3rd ed., Prentice-Hall, Inc., Englewood Cliffs, N.J., 1984.
25. D. E. Johnston, *Printed Circuit Fabrication* **14**, 20 (May 1991); B. A. Anderson, *Printed Circuit Fabrication* **14**, 26 (May 1991).
26. **U.S. Pat. 2,999,770** (Sept. 12, 1961), G. Gutzeit and A. Krieg (to General American Transportation Corp.).
27. *Metal Finishing Guidebook & Directory*, Metals and Plastics Publications, Inc., Hackensack, N.J., 1993.
28. *The Engineering Properties of Electroless Nickel Deposits*, International Nickel Co., New York.
29. A. J. Gould, *Transactions IMF* **66**, 58–62 (1988).
30. R. N. Duncan, *Metal Finishing* **87**, 33 (1989).
31. G. Krulik, R. Sidhu, and N. Simpkins, *Circuitree* **6**(6), 10–12 (June 1993).
32. G. B. Hogaboom, Sr., in H. Narcus, *Trans. Electrochem. Soc.* **88**, 389 (1945).
33. R. G. Wedel, *Int. Met. Rev.* **217**, 97 (1977).
34. G. A. Krulik, *Prod. Finish.* **48**, 49 (1983).
35. G. L. Mallory, *Plating* **61**, 1005 (1974).
36. J. Hajdu and G. A. Krulik, *Plating Surf. Finish.* **70**, 42 (1983); G. A. Krulik, *Plating Surf. Finish.* **71**, 56 (1984).
37. ASTM B604-75, American Society of Testing Materials, Philadelphia, Pa., 1975.
38. L. J. Durney, ed., *Electroplating Engineering Handbook*, 4th ed., Van Nostrand Reinhold Co., Inc., New York, 1984.
39. C. F. Coombs, Jr., *Printed Circuits Handbook*, 2nd ed., McGraw-Hill Book Co., Inc., New York, 1979.
40. S. Nakahara and Y. Okinaka, *Acta Metall.* **31**(5), 713 (1983).
41. M. E. Pole-Baker, *Print. Circuits Fab.* **7**, 26 (Dec. 1984); **8**, 103 (Jan. 1985); **8**, 53 (Feb. 1985).
42. D. Esposito, *Printed Circuit Fabrication* **14**, 72–73 (Oct. 1991).
43. C. R. Shipley, Jr., *Plating Surf. Finish.* **71**, 92 (1984).

44. S. Nakahara and Y. Okinaka, *Acta Metall.* **31**, 24 (1984).
45. H. Nakahara, *Printed Circuit Fabrication* **15**(10), 16–19; (Oct. 1992); **15**(11) 18–21 (Nov. 1992).
46. F. Peadstein and R. F. Weightman, *Plating* **61**, 154 (1974).
47. A. A. Marino and co-workers, *J. Electrochem. Soc.* **132**, 68 (1985).
48. **U.S. Pat. 3,671,291** (June 20, 1972), R. G. Miller and R. L. Cavitt (to PPG Corp.).
49. **U.S. Pat. 3,723,078** (Mar. 27, 1973), K. Parker (to General American Transportation Corp.).
50. **U.S. Pat. 3,995,371** (Dec. 7, 1976), T. J. O'Keefe (to the University of Missouri).
51. A. F. Schmeckenbecker, *J. Electrochem. Soc.* **113**, 778 (1966).
52. A. Mayer and D. S. Catlett, *Plating* **65**, 42 (1978).

Gerald A. Krulik
Applied Electroless Concepts, Inc.



ELECTROLYTIC MACHINING METHODS.

See MACHINING METHODS; ELECTROCHEMICAL.

ELECTROMAGNETIC INTERFERENCE (EMI).

See COATINGS.

ELECTROMIGRATION.

See ELECTROSEPARATIONS.

ELECTRONIC-GRADE CHEMICALS.

See ELECTRONIC MATERIALS; FINE CHEMICALS.

ELECTRONIC MATERIALS

Electronic materials are those exquisitely pure crystal structures which form the basis for the information technology of the 1990s. Since the original theoretical understanding, the progress of solid-state electronics has been paced by the ability to control chemical bonding structures, particularly at surfaces and interfaces (1). A series of chemical discoveries and insights in Ge, Si, and the Group 13–15 (III–V) semiconductors is responsible for the solid-state electronics revolution. Every new device and capability can be identified with a particular chemical breakthrough. The chemistry and chemical engineering required for making electronic materials, ie, these very specific inorganic chemical bonding structures, has come about by techniques employing chemical vapor deposition to effect materials growth and through an increased control over surface and interfacial chemistry (see ELECTRONIC COATINGS; SURFACE AND INTERFACE ANALYSIS; THIN FILMS).

Semiconductor Energy Levels

Semiconductor materials are rather unique and exceptional substances (see SEMICONDUCTORS). The entire semiconductor crystal is one giant covalent molecule. In benzene molecules, the electron wave functions that describe probability density are spread over the six ring-carbon atoms; in a large dye molecule, an electron might be delocalized over a series of rings, but in semiconductors, the electron wave-functions are delocalized, in principle, over an entire macroscopic crystal. Because of the size of these wave functions, no single atom can have much effect on the electron energies, ie, the electronic excitations in semiconductors are delocalized.

Good semiconductors are drawn from the central columns, Groups 13, 14, and 15 (III,IV, and V), of the Periodic Table, where the atoms tend to be nonpolar. For this reason, and because of the giant size of the wave functions, the electron–atom interaction is very weak. The electrons move as if in free space, colliding with the atomic lattice rather infrequently.

In a semiconductor the available energy levels are the valence and conduction bands, which are generally filled and empty, bonding and antibonding, respectively, separated by a forbidden gap as shown in Figure 1. The electrons in the conduction and valence bands act as two separate subsystems. Not only do the electrons ignore the crystallographic lattice of atoms, the electrons in one band tend to ignore those in the other subsystem. This property is unique to electronic materials.

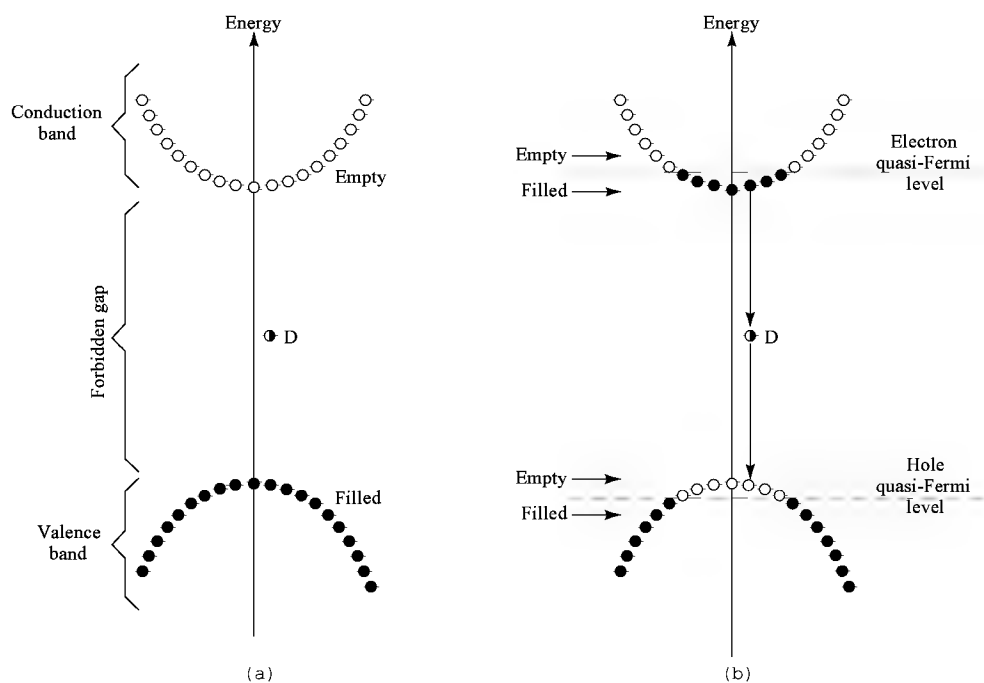


Fig. 1. The energy levels in a semiconductor. Shown are the valence and conduction bands and the forbidden gap in between where ● represents an occupied level, ie, electrons are present; ○, an unoccupied level; and ◐ an energy level arising from a chemical defect D and occurring within the forbidden gap. The electrons in each band are somewhat independent. (a) A cold semiconductor in pitch darkness where the valence band levels are filled and conduction band levels are empty. (b) The same semiconductor exposed to intense light or some other form of excitation showing the quasi-Fermi level for each band. The energy levels are occupied up to the available voltage for that band. There is a population inversion between conduction and valence bands which can lead to optical gain and possible lasing. Conversely, the chemical potential difference between the quasi-Fermi levels can be connected as the output voltage of a solar cell. Equilibrium is reestablished by stepwise recombination at the defect levels D within the forbidden gap.

Electrons excited into the conduction band tend to stay in the conduction band, returning only slowly to the valence band. The corresponding missing electrons in the valence band are called holes. Holes tend to remain in the valence band. The conduction band electrons can establish an equilibrium at a defined chemical potential, and electrons in the valence band can have an equilibrium at a second, different chemical potential. Chemical potential can be regarded as a sort of available voltage from that subsystem. Instead of having one single chemical potential, ie, a Fermi level, for all the electrons in the material, the possibility exists for two separate quasi-Fermi levels in the same crystal.

The idea of having two distinct quasi-Fermi levels or chemical potentials within the same volume of material, first emphasized by Shockley (1), has deeper implications than the somewhat similar concept of two distinct effective temperatures in the same block of material. The latter can occur, for example, when nuclear spins are weakly coupled to atomic motion (see MAGNETIC SPIN RESONANCE). Quasi-Fermi level separations are often labeled as Im_{fer} , Fermi's name spelled backwards.

In the quantum level structures illustrated in Figure 1b, electrons in each band fill up the energy levels, up to the available chemical potential. From Figure 1b two critical optoelectronic devices can be explained.

- (1) Solar cells: the difference between conduction and valence band chemical potentials is the available output voltage of a solar cell. Light creates the chemical potential difference simply by boosting a population of electrons from the valence band into the conduction band (see PHOTOVOLTAIC CELLS; SOLAR ENERGY).
- (2) Lasers (qv): levels just above the valence band chemical potential are essentially (2) empty and unfilled, but levels just below the conduction band chemical potential are filled, permitting a population inversion. Filled levels above, empty levels below, is the principle by which lasers operate (see also LIGHT GENERATION, SEMICONDUCTOR LASERS).

These two devices are the inverse of one another: the semiconductor laser converts electricity into light; the solar cell converts light into electricity.

The possibility of two separate electronic equilibria, ie, the establishment of two quasi-Fermi levels or two different chemical potentials, requires a very slow decay of electrons from the conduction band back into the valence band. The very weak electron–atom coupling, resulting from the large delocalized wave functions in nonpolar materials, permits the slow decay. Electron-hole recombination requires getting rid of the electronic band gap energy, which is usually around one volt, and dumping it off as heat of atomic motion. In relation to characteristic energies of atomic motion, one volt is a huge amount of energy to dissipate in a single step. The weak electron–atom coupling makes nonradiative decay an extremely unlikely event and the low probability for semiconductor materials to dissipate electronic energy as heat is probably their most unique property. By contrast, in organic molecules, decay by nonradiative recombination is sufficiently likely to occur that it is given the name internal conversion.

In fact, nonradiative recombination does occur in semiconductors, but primarily as a result of chemical defects that introduce new energy levels into the forbidden gap. These defect levels act as stepping stones, permitting conduction electrons to cascade down to the valence band in two smaller steps rather than one improbable leap. An energy level within the forbidden gap owing to a defect D is shown in Figure 1. The defects act as electron–hole recombination catalysts, promoting recombination without being consumed themselves. This model of stepwise nonradiative recombination through impurities or other defects is associated with the names of Hall (3) and Shockley and Read (4).

Thus a principal goal of semiconductor materials science has been to create chemically perfect semiconductor structures. Any defects that disturb the perfect valence bonding structure, allowing energy levels in the forbidden gap, must be eliminated as far as possible. Even the utmost extrinsic chemical purity is insufficient, however, because intrinsic defects such as broken bonds, self-interstitials, and vacancies are also proscribed. In particular, unsaturated chemical bonds on the material surface, or in the bulk, contribute nonbonding orbitals having unwanted energy levels in the forbidden gap. The rigid, tetrahedrally coordinated semiconductor crystal structures of silicon [7440-21-3], Si; germanium [7440-56-4], Ge; and gallium arsenide [1303-00-0], GaAs, have a tendency to reject both extrinsic and intrinsic defects, contributing to their technological success.

Semiconductor Surfaces

Semiconductor surfaces are the most likely location for intrinsic defects such as dangling or weak bonds to occur. The need for surfaces of extremely high chemical quality can be appreciated from examination of Figure 2 showing the point contact germanium transistor. The emitter point contact emits a cloud of minority carriers (Fig. 2a), which by definition are the least dense of the two types of carriers, either electrons or holes. As indicated in Figure 2b, these minority carriers undergo a random walk to find the small-area collector point contact. If the transistor is to have any significant gain, almost all the minority carriers must be collected. During the random walk these carriers must pass a gauntlet of defects any one of which could catalyze recombination

(see INTEGRATED CIRCUITS).

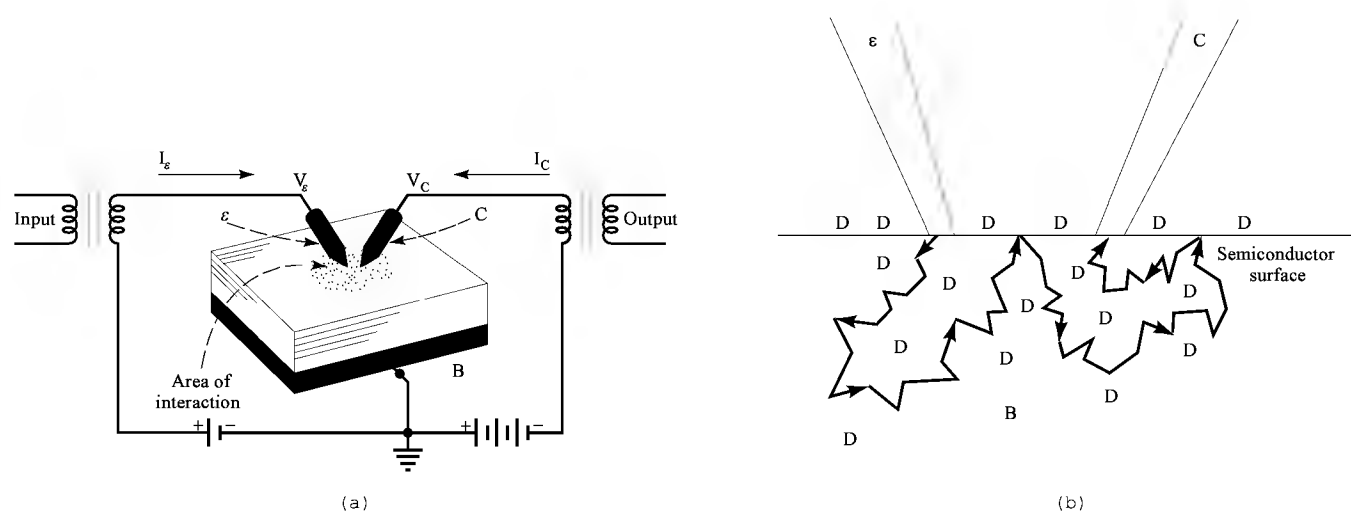


Fig. 2. A germanium transistor where B is the base, C, the collector, and ε, the emitter. (a) The external transistor circuit showing the diffusion cloud of minority carriers around the emitter point contact; I_C and I_ϵ , and V_C and V_ϵ , are the current and the voltage of the collector and emitter, respectively. (b) The area of interaction showing a typical random walk the minority carriers must undergo, passing a gauntlet of defects D in order to find the collector point contact. In practice, the surface concentration of defective chemical bonds must be less than 1 per 10,000 surface bonds for the carriers to have any significant probability of finding the collector before recombining. Defects are usually less than 1 in 10^4 to 1 in 10^7 at the surface and less than 1 in 10^4 to 1 in 10^{11} in the bulk material. For the transistor to have any significant gain, almost all the carriers must find the collector.

On average, a carrier scatters off the surface thousands of times before it finally stumbles its way into the collector. If even a tiny fraction of the surface were covered with defects, the carriers would never make it to the collector. A germanium surface covered with a natural air-grown oxide has less than 1 part per 10,000 of defective chemical bonds on the surface. The discovery of the Ge point-contact transistor was therefore critically dependent on this chemical property. Indeed the point-contact transistor is not possible using silicon or any of the other semiconductor compounds known as of this writing. The density of chemical defects at interfaces having air-grown oxides is generally too high.

The bulk chemical defect densities that can be tolerated in solid-state electronics range from 1 in 10^4 to 1 in 10^{11} , depending on the specific application. Corresponding surface defect densities that can be tolerated range from 1 in 10^4 to 1 in 10^7 , ie, nearly all of the semiconductor surface atoms must be cleanly saturated with strong covalent bonds, because defects introduce energy levels into the forbidden gap. These requirements for semiconductor surfaces and interfaces give rise to a chemical figure of merit, ie, equivalent to a surface chemical reaction having a 99.9999% to 99.999999% yield. The surface of germanium minimally satisfies this requirement in air, permitting the discovery of the first transistor, but the germanium surface cannot be further improved.

The Role of Silicon

The Si–SiO₂ Interface. Beginning in the mid-1950s (5), thermal oxidation of silicon at high temperatures in oxygen was begun, coating the silicon with a thin layer of silicon dioxide [7631-86-9], SiO₂, glass. The thermal oxidation recipe was gradually perfected (6,7) and by the late 1960s the figure of merit for the Si–SiO₂ interface had been improved to 1 defective chemical bond in 10^8 . The Si–SiO₂ interface is an amorphous crystalline heterojunction. The interfacial bonds can be 99.9999% saturated. Thus, in short order the microprocessor (1969), the memory chip (1971), and the pocket calculator (1972), were developed. In 1975 *Time* magazine declared the microcomputer to be the Man-of-the-Year. In 1992, microchip production was a $\sim \$6 \times 10^{10}$ annual industry worldwide, supporting a $\sim \$5 \times 10^{11}$ systems, software, and communications industry and millions of highly educated people.

Purification of Silicon. Chemical purity plays an equally important role in the bulk of materials as on the surface. To approach the goal of absolute structural perfection and chemical purity, semiconductor Si is purified by the distillation of trichlorosilane [10025-78-2], SiHCl₃, followed by chemical vapor deposition (CVD) of bulk polycrystalline silicon.



Purified polycrystalline CVD silicon from this reaction is then melted and a single-crystal boule weighing as much or more than 50 kg, and having a diameter up to 20 cm, is pulled from the melt by Czochralski growth (8). Metallurgical-grade silicon is not sufficiently pure for applications in electronics (see SILICON AND SILICON ALLOYS).

The distillation of SiHCl₃ is a relatively straightforward and low cost purification method. On the other hand, the reprecipitation of elemental silicon by CVD is expensive and complicated because of the need to maintain the utmost purity at highly elevated temperatures. Long narrow rods of polycrystalline silicon are heated by an electric current to $\sim 1100^\circ\text{C}$ providing the thermodynamic conditions for silicon precipitation. Because the yield in equation 1 is only $\sim 10\%$, there is considerable recycling of chlorosilane reagents. In this respect, the semiconductor silicon factory is typical of most chemical plants. The growth of the polycrystalline silicon rod is diffusion rate limited, $\sim 1 \mu\text{m min}$, taking as long as two weeks to grow out to a 10-cm diameter. Therefore the process is very slow, very capital intensive, and because heat is continually radiated away at 1100°C during the long growth period, very energy intensive as well. Heat radiation reflectors are employed to reduce energy cost, as well as clusters of rods which intercept one another's thermal radiation. Owing to the electrical energy cost component, semiconductor silicon plants are frequently located in regions having low cost hydroelectric power, such as the Pacific Northwest of the United States. To further reduce the energy cost, silane [7803-62-5], SiH₄, gas decomposition is being used in at least one large plant to produce semiconductor silicon in large volumes. The advantage of using silane is that decomposition takes place at a diminished temperature, $\sim 800^\circ\text{C}$.

Semiconductor-grade polycrystalline silicon costs about \$40/kg. Of the two high temperature processing steps, CVD and crystal growth, the CVD process is more responsible for the high monocrystalline silicon boule costs. The single-crystal boule that finally results from CVD and crystal growth can have a bulk purity of 1 part in 10^{11} , not counting oxygen and carbon impurities which are relatively benign. Normally the entire boule is structurally perfect, ie, there are no dislocations. Worldwide, around 40,000 metric tons of electronically pure silicon are produced per year and the cost is less than \$10 per 10-cm diameter wafer.

III–V Semiconductors

For optoelectronics the binary compound semiconductors drawn from Groups 13 and 15 (III and V) of the Periodic Table are essential. These often have direct rather than indirect band gaps, which means that, unlike Si and Ge, the lowest lying absorption levels interact strongly with light. The basic devices of

optical communications, light-emitting diodes (LEDs) (see LIGHT GENERATION, LIGHT EMITTING DIODES) and semiconductor lasers, are made of these III–V semiconductors. Aluminum arsenide [22831-42-1], AlAs, GaAs, and the alloys of these compounds have historically been the most important III–V material system. The reason once again derives from the need to control interfacial chemical bonding structures.

The double heterostructure, invented in the early 1960s (9), is shown schematically in Figure 3. It is essentially a crystalline sandwich: the bread is made of AlAs and the filling of GaAs. Because the band gaps of AlAs and GaAs are 2.2 eV and 1.4 eV, respectively, the GaAs wave functions are sandwiched in by the 2.2 eV potential barriers, as shown. Although the electrons and holes are prevented from seeing any external surface, they do see the AlAs–GaAs interface. However, the cubic unit cell dimensions of GaAs and AlAs are 0.56533 nm and 0.56605 nm, respectively. Thus the mismatch at the interface is less than 0.1% meaning that the crystal structures can match up nearly perfectly, leaving only a few unsaturated chemical bonds at the interface. Generally, the interfacial bonds in the Ga_{1-x}Al_xAs system are 99.999% saturated. Although this is not as good as the best Si–SiO₂ interfaces, it is excellent nonetheless. The growth of successive atomic layers of semiconductor material, in perfect registry with atoms in the underlying crystal, is called epitaxy; the perfect atomic registry between layers of differing composition is called heteroepitaxy.

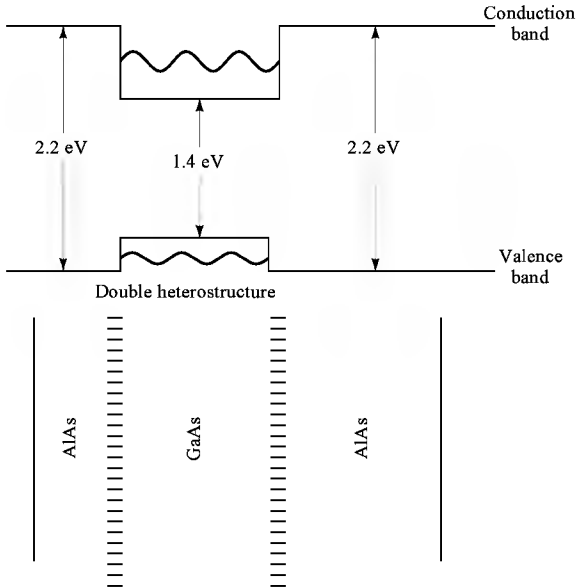
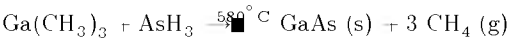


Fig. 3. The lattice-matched double heterostructure, where the waves shown in the conduction band and the valence band are wave functions, $\Psi(x)$, representing probability density distributions of carriers confined by the barriers. The chemical bonds, shown as short horizontal stripes at the AlAs–GaAs interfaces, match up almost perfectly. The wave functions, sandwiched in by the 2.2 eV potential barrier of AlAs, never see the defective bonds of an external surface. When the GaAs layer is made so narrow that a single wave barely fits into the allotted space, the potential well is called a quantum well. Because of the match in the atomic spacings between GaAs and AlAs, 99.999% of the interfacial chemical bonds are saturated.

When the sandwich in Figure 3 is made very narrow, quantum confinement influences the wave functions in the GaAs. Excellent control is possible because the layers can be grown one atomic monolayer at a time (10). The wave functions can also be modified by alloying aluminum [7429-90-5], Al, or indium [7440-74-6], In, into the GaAs. Moreover, arsenic [7440-38-2], As, can be replaced by phosphorus [7723-14-0], P, or antimony [7440-36-0], Sb. Furthermore, lattice matching can be accomplished by intentionally straining mismatched systems (11). Much research is oriented toward extending these artificial structures into the third dimension (12).

Inorganic semiconductor crystals have innumerable possible arrangements. Whereas there has been a long-standing call (13) for a molecular electronics based on organic molecular structures, inorganic quantum structures are already being made and there is a well-developed inorganic electronics. Hydrocarbon-based molecules do not satisfy the requirements of wave function delocalization, weak electron–atom coupling, and structural perfection to the same degree that the successful inorganic semiconductors do.

Physics and chemistry researchers approach III–V synthesis and epitaxial growth, ie, growth in perfect registry with the atoms of an underlying crystal, differently. The physics approach, known as molecular beam epitaxy (MBE), is essentially the evaporation (14–16) of the elements, as illustrated in Figure 4. The chemistry approach, organometallic chemical vapor deposition (OMCVD) (17) is exemplified by the typical chemical reaction:



Thin-film epitaxy by OMCVD is generally more flexible, faster, lower in cost, and more suited for industrial production than MBE. An OMCVD system usually consists of two principal components, a gas manifold for blending the gas composition, and a graphite substrate holder which is usually inductively heated. A schematic diagram of an OMCVD system is shown in Figure 5.

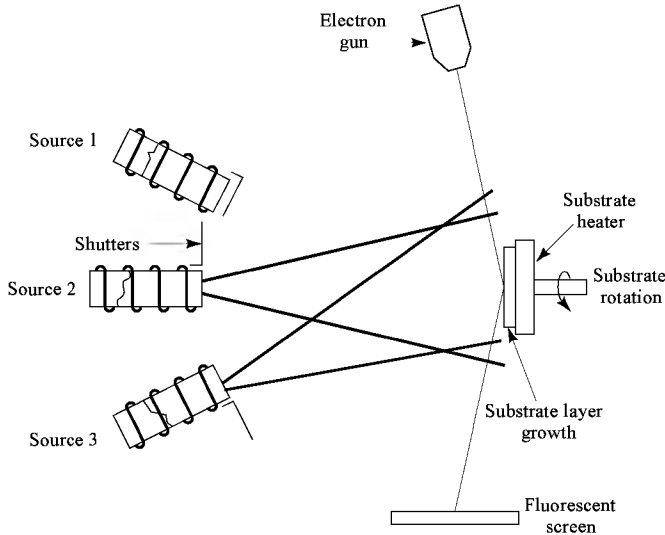


Fig. 4. Schematic of an ultrahigh vacuum molecular beam epitaxy (MBE) growth chamber, showing the source ovens from which the Group III–V elements are evaporated; the shutters corresponding to the required elements, such as that in front of Source 1, which control the composition of the grown layer; an electron gun which produces a beam for reflection high energy electron diffraction (rheed) and monitors the crystal structure of the growing layer; and the substrate holder which rotates to provide more uniformity in the deposited film. After Ref. 14, see text.

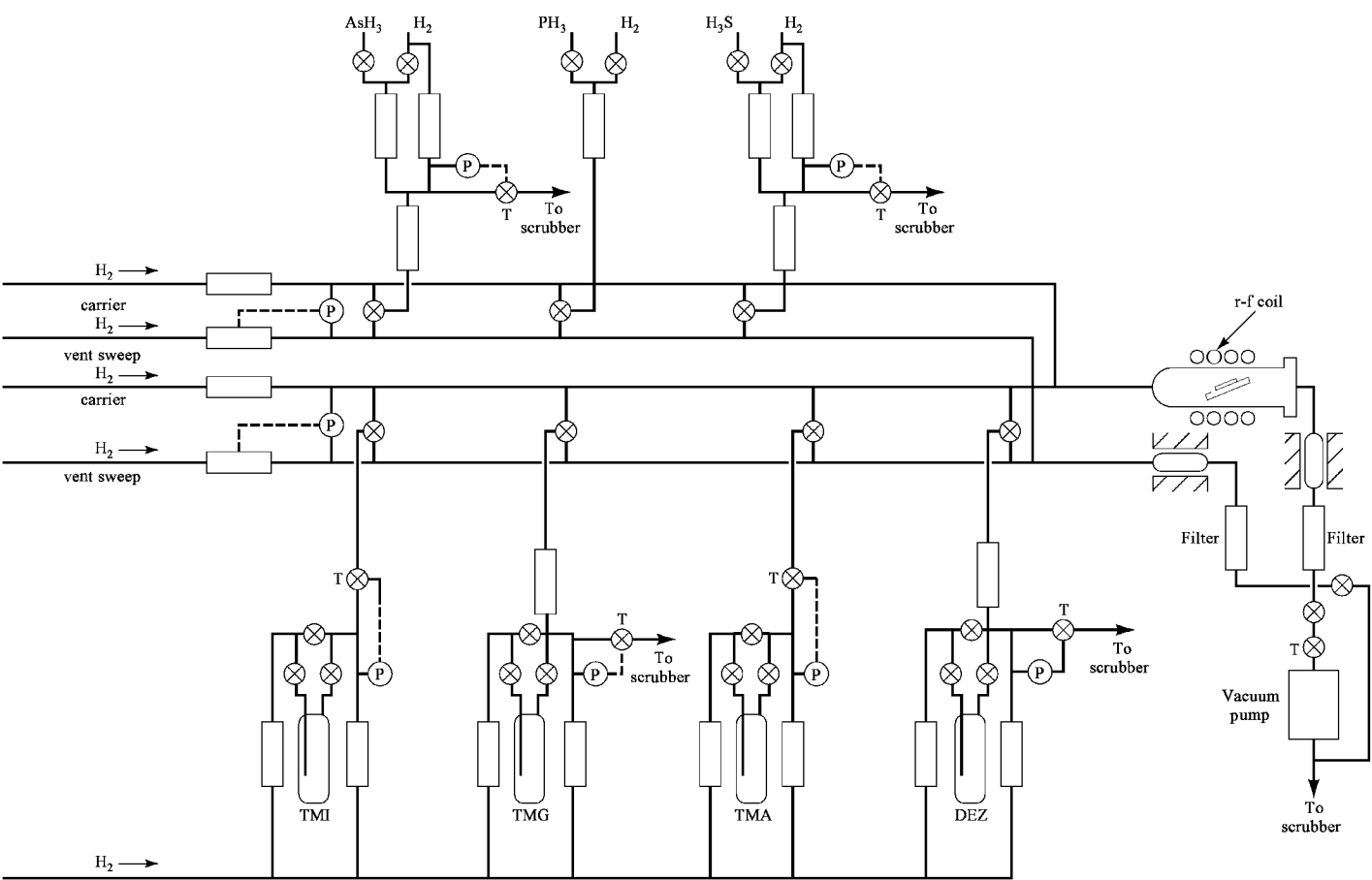


Fig. 5. A gas manifold for the production of epitaxial layers of Group III–V semiconductors by OMCVD where  represents a mass flow controller; , a pneumatic valve; , a throttle valve; , pressure sensor and controller; and , a pyrolysis oven. Reactive gases are swept along by ultrapure hydrogen gas. Gases are trimethylindium [3385-78-2] (TMI); trimethylgallium [1445-79-0] (TMG); trimethylaluminum [75-24-1] (TMA); and diethylzinc [557-20-0] (DEZ). The semiconductor wafer is on a graphite susceptor inside the radio frequency (r-f) coil (17).

The development of double-heterostructure III–V semiconductor lasers (18) together with the demonstration of low loss SiO₂ glass fibers (19) which both occurred coincidentally in 1969, gave birth to optical communications. Optical fibers are the technology of choice in newer trunk and long-distance communications installations, and optical fibers are expected to connect to individual telephone subscribers soon, making unprecedented bandwidth available for revolutionary new services (see FIBER OPTICS). The materials for these fibers are prepared by these thin-film crystal growth processes, MBE and OMCVD.

In an MBE growth reactor, the Group III–V elements are evaporated from source ovens in a growth chamber having background pressure in the 10⁻⁹–10⁻⁸ Pa (10⁻¹¹–10⁻¹⁰ torr) range. The crucible holding the Group III–V elements is usually made of boron nitride [10043-11-5], BN, or alumina [1344-28-1], Al₂O₃, which is radiatively heated by a refractory metal filament (see REFRACTORIES). The composition of the growing layer is controlled by shutters, which correspond to the respective elements and source ovens. The crystal structure of the growing layer is monitored by reflection high energy electron diffraction (rheed).

In contrast to MBE, OMCVD can be carried out either at atmospheric or at low pressures. A critical role is played by ultrapure hydrogen [1333-74-0], H₂, which is produced by diffusion through a metal filter made of a heated palladium [7440-05-3], Pd, membrane. Only hydrogen gas can diffuse through such a membrane; thus, chemical purity is maintained even at high pressures. Substrates are placed on a graphite susceptor and heated by either radio frequency induction or infrared radiation. The gas manifold shown in Figure 5 is designed to allow abrupt changes in gas composition, corresponding to abrupt interfaces between layers of different composition in the growing material. The ability to select various precursor gases (20) provides considerable flexibility for optimizing the reactor conditions. For example, some gases are considerably less toxic than the arsine [7784-42-1], AsH₃, and phosphine [7803-51-2], PH₃, which have traditionally been used for OMCVD. Other gases can produce less carbon contamination, or conversely can produce more efficient doping by carbon, etc. The gas composition parameters are likely to be an important engineering variable as OMCVD develops in the future.

Control over the inorganic chemistry of semiconductor crystals, surfaces, and interfaces has determined the rate of progress in solid-state electronics. Whereas inorganic electronic materials have been emphasized herein, organic materials are also important to the electronics revolution, not as the active electronic material, but as enabling components. For example, photoresists are photographic materials that allow the patterning of microcircuits (see LITHOGRAPHY; PHOTOCONDUCTIVE POLYMERS), and insulators, such as the polyimides (qv), are helpful in microcircuit fabrication and packaging (see INSULATION, ELECTRIC; PACKAGING, SEMICONDUCTOR AND ELECTRONIC MATERIALS).

BIBLIOGRAPHY

1. W. Shockley, *Electrons and Holes in Semiconductors*, Van Nostrand, New York, 1950.
2. E. Yablonovitch and E. O. Kane, *J. Lightwave Tech.* **LT-4**, 504 (1986). In present-day semiconductor lasers, E_{F_i} does not extend all the way down into the valence band as shown in Figure 1, penalizing efficiency.
3. R. N. Hall, *Phys. Rev.* **87**, 387 (1952).

4.

W. Shockley and W. T. Read, *Phys. Rev.* **87**, 835 (1952).

5.

M. M. Atalla, E. Tannenbaum, and E. J. Scheibner, *Bell Syst. Tech. J.* **38**, 749 (1959).

6.

B. E. Deal, E. H. Snow, and A. S. Grove, *Semicond. Prod. Solid-State Technol.* **9**, 25 (1966).

7.

A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley & Sons, Inc., New York, 1967.

8.

K. E. Benson in S. Mahajan and L. C. Kimerling eds., *Encyclopedia of Semiconducting Materials and Related Technologies*, Pergamon Press, Oxford, UK, 1992, p. 467.

9.

H. Kroemer, *Proc. IEEE* **51**, 1782 (1963); **USSR Pat. 181737** (1963). Z. I. Alferov and R. F. Kazarinov.

10.

M. A. Tischler, N. G. Anderson, and S. M. Bedair, *Appl. Phys. Lett.* **49**, 1199 (1986).

11.

G. C. Osbourn, R. M. Biefeld, and P. L. Gourley, *Appl. Phys. Lett.* **41**, 172 (1982).

12.

M. Tsuchiya and co-workers, *Phys. Rev. Lett.* **62**, 466 (1989).

13.

F. L. Carter, ed., *Molecular Electronic Devices*, Marcel Dekker, Inc., New York, 1982.

14.

J. P. Harbison in Ref. 8, p. 306.

15.

L. L. Chang and K. Ploog, eds., *Molecular Beam Epitaxy and Heterostructures*, Martinus Nijhoff, Dordrecht, The Netherlands, 1985.

16.

M. A. Herman and H. Sitter, *Molecular Beam Epitaxy: Fundamentals and Current Status*, Springer-Verlag, Berlin, 1989.

17.

R. Bhat and S. Mahajan in Ref. 8, p. 349.

18.

M. B. Panish and I. Hayashi, *Appl. Phys. Lett.* **16**, 326 (1970).

19.

R. D. Mauer, *Proc. IEEE* **61**, 452 (1973).

20.

G. B. Stringfellow, *OMVPE: Theory and Practice*, Academic Press, Boston, Mass., 1989

Eli Yablonovitch
University of California at Los Angeles

ELECTRONICS, COATINGS

Coating technology can be defined as replacing air at the surface of a substrate to give a film structure having varying layers of different properties. This technology covers a wide range of products and processes (see COATING PROCESSES; COATINGS; PAINT). Electronics coatings refers to the wide variety of coated structures used in electronic devices.

There have been significant advances in electronics coatings technology in the latter part of the twentieth century, leading to rapid growth and high economic importance. These coatings have some unique aspects of formulation, production, and use. The ability to produce high resolution coatings, at high volume and low cost, has been the driving force for the revolutionary advances in a wide variety of electronic components, integrated circuits (qv), memory chips, magnetic storage devices, cathode ray tubes, optical storage disks, etc (see also ELECTRONIC MATERIALS; INFORMATION STORAGE MATERIALS). Improved functionality of these components has led to advances in a wide range of consumer, industrial, and military products. Typical devices are television sets, video cassette recorders, cameras, fax machines, cellular phones, personal and lap-top computers, fly-by-wire control systems for airplanes, and the guidance system for such military ordnance as smart bombs and cruise missiles.

Significant advances have also been made in the more traditional area of wire and cable used to transmit the electricity needed for these devices (see ELECTRICAL CONNECTORS). The optical cables (see FIBER OPTICS; NONLINEAR OPTICAL MATERIALS) that are replacing the traditional copper (qv) cables for telephone usage require very sophisticated coatings to protect cable from light loss and deterioration. The ribbon cable and multiple coaxial cable assemblies used for cable television, stereo systems, and burglar alarms all depend on the cable coatings reducing interference with each other so that signal quality is maintained.

Economic Aspects

Electronic coatings are of significant economic importance, as are the finished products in which they go. The worldwide total value of the resulting products is \$500 billion. Table 1 provides a geographic breakdown. The annual electronics coatings market value is estimated to be \$5 billion. These coatings are manufactured in several countries. Some of the principal manufacturers of electronic coatings are

Company	Products
Minnesota Mining and Manufacturing W. R. Grace & Co.	conformal coatings dielectric and conductive adhesives, encapsulants, heat dissipating materials, manufacturing aid coatings
Union Carbide Corp.	Parylene conformal coatings, photoresists, developers, etchants, solder masks, potting compounds
Conap unit of American Cyanamid Du Pont Co.	potting compounds, conformal coatings conformal coatings, photoresists, manufacturing aids, electronic functional coating compounds
OGG Microelectronic Materials, Olin CIBA-GEIGY venture Dexter	photosensitive polyimides, high performance semiconductors encapsulants
Dow Corning	potting and encapsulating compounds
Asahi	packing materials
Nippon Kayaku	packing materials
General Electric	packaging materials
Unilver	adhesives, dielectric coatings, protective coatings, dielectric interlayers, electronic functional coatings

Table 1. Worldwide 1992 Electronics Industry Value,^a \$ × 10⁹

Market	Geographic area		
	United States	Japan	Europe
computers and office equipment	103	54	119
communications	36	12	36
consumer	35	20	32
semiconductors	17	24	11
capital and test equipment	10	6	4
Total	519		

^a Ref. 1.

Functions of Electronic Coatings

Electronics coatings can generally be classified as: (1) functional, ie, coatings that provide a variety of electrical, magnetic, optical, chemical, mechanical, and or thermal properties enabling a device to function as prescribed; (2) manufacturing aids, ie, coatings that serve as an integral part of the manufacturing process to make a specific component or device; or (3) protective decorative, ie, coatings that protect the component or device from environmental damage during its normal use cycle. As can be seen from Table 2, most electronic devices require coatings that perform more than one of these functions. A wide variety of chemical compounds or formulations are needed to meet these specifications.

Table 2. Coating Requirements of Electronic Products^a

Device	Coating function ^b
--------	-------------------------------

optical fibers	P,F
integrated optics	M,F,P
wires	P,F
printed circuit boards	M,F,P
passive components ^c	F
integrated circuits	M,P,F
liquid crystal displays	M
lenses	M,F
cathode ray tubes	M,P
projection TV screens	F
solid-state cameras	M,F
compact disks	M,P,F
LaserVision disks	M,F
optical recording	M,F,P
magnetic recording	M,F,P

^a Ref. 2.
^b F = functional, ie, electrical, magnetic, thermal, mechanical, chemical, or optical function; M = manufacturing aid; and P = protective.
^c Components such as capacitors and resistors.

Functional Coatings. Whereas there are many types of functional coatings, the basic properties or characteristics that give a device utility are its electrical, magnetic, and optical properties. The electrical properties can be defined by the ability of the material to conduct or hinder flow of electrons. This characteristic resistance is expressed by

$$R = \frac{\rho L}{A}$$

where *R* = resistance of conductor in ohms, *L* = length in cm, *ρ* = proportionality constant called specific resistance or resistivity in ohm·cm, and *A* = area in cm². The resistivity is typically expressed in terms of the volume resistivity, ie, the ohmic resistance of a cube of material, having dimensions of 1 cm per side. This is expressed as ohm·cm. Good conductors, which do not impede the flow of electrons, have values of 10⁻⁶–10⁻⁸ ohm·cm. Poor conductors impede the flow of electrons and have values greater than 10⁸ ohm·cm. Figure 1 shows the comparative resistivity of a variety of materials.

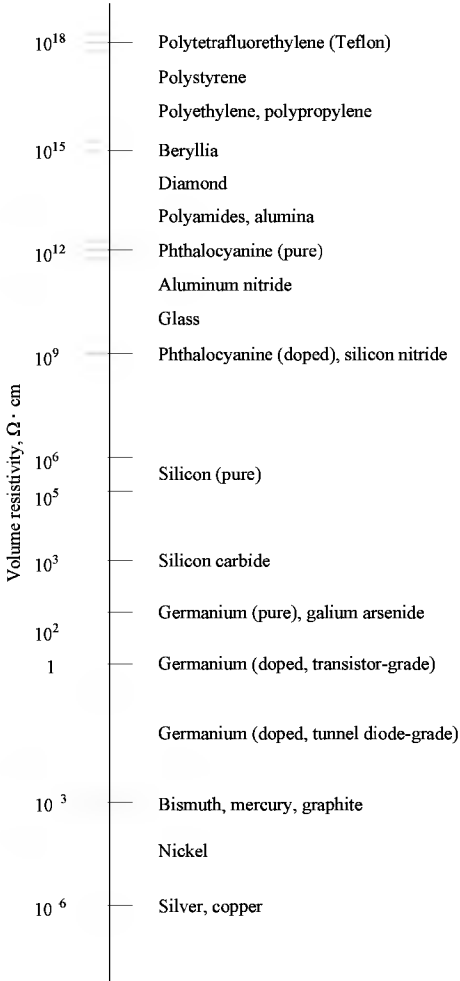


Fig. 1. Electrical resistance materials.

Resistivity can be used as a guide to the role a material performs in a specific device. Materials having high values, such as Teflon, serve an insulation function (see INSULATION, ELECTRICAL). Metals such as silver and copper are excellent conductors. Organic compounds and polymers can cover a wide range of values, and the actual resistivity depends on exact composition.

One factor leading to the growth of electronics coatings technology has been the ability to modify compounds to give very specific electrical properties. Variations in compositions and additions of trace materials can have a significant effect on resistivity, particularly in the chemistry of semiconductors (qv), where trace amounts of dopants can significantly change resistivity and thus the function. Polymers can be compounded with metals to increase the polymer conductivity while utilizing the polymer mechanical properties to give flexibility in fabrication of specific elements. The exact conductivity is, of course, a function of the dopant used. Polymers can also be custom synthesized to give increased conductivity. Typically, polymers having a high degree of conjugation and linear backbones tend to be better conductors than others.

Optical functions are important both to transmit information through fiber optic cables and to store information on optical disks. An optical fiber

consists of a central core, such as silica [7631-86-9] surrounded by a series of coatings to protect the fiber and give it flexibility. Typically uv-curable acrylates are used for these protective coatings (see ACRYLIC ESTER POLYMERS). To minimize light loss, the silica is doped with oxides of germanium or phosphorus to raise the refractive index.

Optical information storage is a relatively new method for storing large quantities of information. The compact disk (CD) and LaserVision are two well-known read-only formats that have been on the market since the early 1980s, whereas write-once and erasable (rewritable) recordable formats have been introduced more recently. Photopolymerizable acrylate compositions are used for abrasion protection of the metallic mirrors of CDs, as the information carrier in the replication of LaserVision video disks, and in the replication of aspheric lenses used in the optical heads of the drives or players. Numerous coatings have been developed for direct writing of the digital code onto recordable disks. These may be polymer dye, metallic semiconductor alloy, or even magneto-optic in composition. Writing occurs by a thermally induced optical change in the thin recording layer by a focused laser beam. Readout is typically done in reflection at lower laser power.

Because there are many other properties that also are important, coatings cannot be selected only on this basis. The mechanical and chemical properties of the coating, change of properties with temperature, dielectric and adhesion properties, and particularly the cost of fabrication are all important parameters. Coatings can also be used to transport heat created away from a component and keep the component functioning as designed, or to protect a component from temperature variations in the environment.

Manufacturing Aids. The use of coatings to aid in the manufacture of an electronic component is somewhat unique to electronics coatings. Most coatings, once applied, are part of the finished product, and remain with the coated structure throughout the majority of its life cycle. The manufacturing aid coating, however, is only temporary, and is used as part of the manufacturing process to help apply functional coatings in a desired path or to locate specific functions to give the component its desired characteristics. Historically, tubes, resistors, capacitors, etc, were placed on a board and connected using wires soldered in place. This methodology could not produce conductor paths below the 300 μm width needed for high performance devices. As a result, it has been replaced by processes that use temporary coatings, such as resists, to produce the conductor lines. The use of manufacturing aid coatings also allows manufacturers to rapidly and precisely apply a wide variety of newer types of coated materials that perform the same resistance and capacitance function previously done using discrete components. Electronics coatings technology can easily produce conductor paths of 1–5 μm , and even 0.35 μm is possible. The processes are easily automated, and have led to improvements in the quality, costs, and long-term performance of the resulting circuits.

Using electronics coatings technology, the temporary coating is applied to an underlayer or substrate, and a line or pattern copied to the temporary coating using uv light, x-rays, or an electron beam (see LITHOGRAPHY). The original pattern is prepared using computer design programs, and then transferred to photographic film (see PHOTOGRAPHY). The copying process can also significantly reduce the size of the pattern, so that the circuit can be designed in large scale and then reduced for the miniaturized part. The significant reduction in the size of lines produced yields an increased efficiency of all components.

The patterned coating leaves part of the underlayer protected and part unprotected. Conductive, resistive, or capacitive coatings can then be applied as needed to the unprotected portion to give functionality and conductor lines where desired. The temporary coating is then removed, and the process may be repeated using additional temporary coatings that give different patterns. This process may be used to apply or add other functions, to solder components in place, etch the underlayer, or apply dopants. When the component is finished, very little of the temporary coating remains.

Typical of the temporary or manufacturing aid coating systems is the RISTON dry film photoresist for printed circuit (PC) board fabrication. This was the first of these systems developed. The RISTON product structure and the basic steps in its use are shown in Figure 2. It consists of a photopolymer sheet laminated between a Mylar cover sheet and a polyolefin separation sheet. It is manufactured as a continuous web (see COATING PROCESSES, SURVEY), and is supplied in rolls of varying width and photopolymer composition.

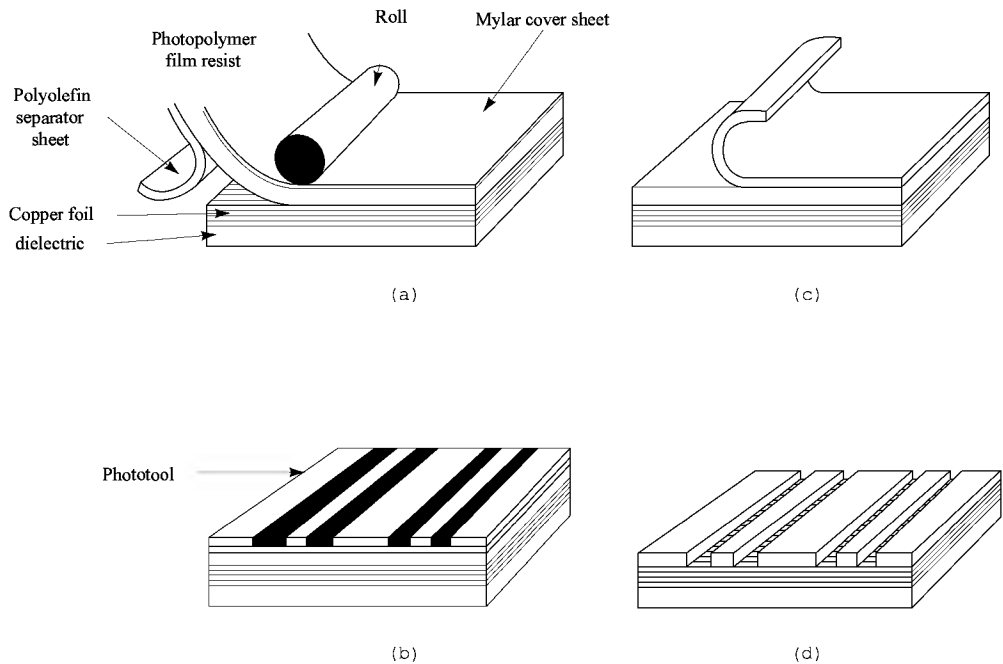


Fig. 2. Schematic of the RISTON dry film photoresist process. (a) Removal of polyolefin separator sheet and laminate resist to clean surface, using special laminator; (b) exposure to uv source using positive or negative phototool (positive to plate; negative for print-and-etch); (c) removal of the protective Mylar, which is readily removed by hand; and (d) development using a special processor (3).

The steps in the process shown in Figure 2 are (1) clean the substrate by mechanical brush scrubbing or by using either spray or dipped solvents, vapor degreasing, or ultrasonic assist (microetchants may also be used); (2) laminate the photopolymer resist film to one or both sides of panel (the separator layer is removed just prior to lamination); (3) expose the board using a phototool through the Mylar cover sheet to harden the photopolymer to give the desired pattern; (4) remove the Mylar cover sheet and wash away the unexposed photopolymer using an alkaline water solution or an appropriate solvent; (5) apply the desired functional electronic coating; and (6) strip the photopolymer resist film from the circuit board, typically also using an aqueous alkaline solution. This process is highly automated and uses specific equipment to perform all of the functions, giving high productivity, low cost, and the precision needed.

In the production of integrated circuits where there is a need for very small ($<1\ \mu\text{m}$) lines, which have to be very close to each other in order to achieve the miniaturization and small circuitry that is required, extreme cleanliness is required in all of the process elements. Dirt, particles, or defects in the temporary coating that are larger than 1 μm can be imaged, and ultimately give an electrical connection between two lines that interfere with the functioning of the circuit. Thus all of the process components must be extraordinarily clean and free from defects and nonuniformities and the goal of all electronics coatings is to be defect-free. This sensitivity to contamination and the requirement for cleanliness are other differentiating elements from most

coating processes, where much higher defect sizes can be tolerated.

To meet the cleanliness need, all elements of the process are controlled to minimize sources of contamination. Air normally contains a large volume of contaminants in the form of dirt, dust, and pollen. The human body sheds a large volume of particulate contaminants such as skin, phlegm, hair, etc. The importance of eliminating these contaminants can be understood by comparing the size of human hair (50 μm diameter) and a bacterium (~5 μm) to a 1-μm line such as used in a memory chip. To keep particulates from contacting the coating, all machines are maintained in clean rooms where air is filtered to remove contaminants; also, people entering the room wear protective clothing that does not shed particulates into the air. The clean rooms are rated by the number of particles present in the air. A class 100 rating indicates 100 or fewer particles >0.5 μm are present in 1 ft³ (2.83 × 10⁻² m³) of air; a class 10,000 would have 10,000 or fewer 0.5 μm particles. The exact specification depends on need, but class 100 rooms are in routine use.

In addition, all of the process raw materials must be clean and not introduce contaminants. The raw materials and temporary coatings must also be defect-free, and these have to be manufactured under similar conditions so that no contaminants are introduced. The solvents used to clean the substrate and develop the resists must be filtered and pure. Care must also be taken to ensure that no trace compounds or elements are present that may affect the electronic properties. The specific type of coating aid, the type of functional coating, and the process used to apply the functional coating are all widely varied in actual practice.

Protective/Decorative Coatings. After the component or finished device has been fabricated, coatings may provide three additionally needed functions: protection of the component from the environment, to ensure the same performance throughout the component's useful lifetime; protection of the user and the environment from the component; and a decorative and characteristic appearance. Examples are the three copper conductor strands of wire in a typical electrical appliance. Each wire strand has a polymeric coating to protect the strands from contacting each other, thus preventing a short circuit and protecting the user when the cord is touched. The coating must also protect the copper from attack from the environment so that the metal does not corrode and fail. The coating should also enhance the appearance of the device. Wire coatings are color coded to ensure correct polarity when devices are attached. Black, red, blue, or yellow are the hot lead, power source; white is neutral; and green is ground.

The requirement for a protective coating for components depends on the environment in which it is going to be used and the potentially destructive capability of the various components present. In normal use, all devices are exposed to air, which contains water vapor and oxygen. Both react with the functional, electronic, magnetic, or optical coatings to change performance. Whereas it is possible to hermetically seal memory chips and integrated circuits during shipment and storage to eliminate these effects, it is not feasible to do this for the vast majority of uses. Thus protective coatings are used that retard the passage of moisture, oxygen, and sulfur dioxide. These are selected on the basis of low permeability and low water absorption as well as on aging performance. Several polymer classes, eg, polyurethanes (see URETHANE POLYMERS), polyimides (qv), poly(vinylidene chloride) (see VINYLIDENE CHLORIDE AND POLY(VINYLIDENE CHLORIDE)), are good for this application. If the device or component is to be used in a special environment such as the ocean, the ability to protect against salt water and corrosion resistance to salt become key factors.

Electronic devices can also generate electromagnetic and radio frequency interference waves that can interfere with other electronic devices. These waves must be modulated and leakage to the environment prevented. Plastics, silicones, acrylics, and polyesters (qv) that are filled with conductive fillers, such as silver, nickel, and copper, are used for this application (1). Although nickel-filled polymers are low cost and efficient, these are not preferred because of the carcinogenic nature of nickel powder.

Coatings Properties

Material property specifications must be written by design and material engineers to control engineering requirements and to control incoming raw material quality. Material property requirements depend on various in-use functional needs in terms of electrical, mechanical, thermal, chemical, optical, and magnetic properties.

Electronic coatings can be classified as either organic, inorganic, or metallic. Organic coatings are typically polymeric in nature. A wide variety of polymer types are used. Although polymeric properties vary considerably less over the full range than those of the inorganics and metallics, both chemistry and molecular structure play a role because these dictate properties and coating behavior. A list of polymer uses in electronic coatings follows. More detail can be found in the literature (4).

Polymer	Uses
alkyd	polyesters for protective and decorative coatings; paints
acrylic	general purpose for molding, casting, and coating formulations; dip and sprayable PC board coatings
epoxy	protective and electrical-grade adhesives, electronic encapsulation, composites, and PC boards; solder masks
fluorocarbons	good chemical and electrical properties, protective coatings, release coatings, barrier coatings
phenolic	work-horse thermoset for adhesives; parts
polyimide	high temperature applications, glass-reinforced PC boards, flexible film and cable, interlayer dielectrics
polyurethane	tough, abrasion resistant for casting, potting, and encapsulation of connectors and modules; conformal coatings for PC assemblies
polyvinyls	chlorides, fluorides, vinylidene chlorides and fluorides, vinyl aldehydes, and polystyrene; used for moisture barriers, primary-wire insulation, corrosion protection, dielectric impregnants, and baking enamels
polyxythylene	conformal coatings for PC assemblies; insulation
silicone	available as solvent solutions, room-temperature vulcanizable (RTV) rubbers, and solventless resins; used as insulation for high temperature and high voltage parts

Components

Printed Circuit Functional Coatings. A printed circuit (PC) is an electrical connection or circuit board assembly fabricated using any of a number of graphic arts processes. The ability to be mass produced, high degree of reliability, and small size and weight, ie, increased circuit density, have made printed circuits the elements of choice in all kinds of electronic devices and equipment. Printed circuit applications can be printed wiring; hybrid circuits, ie, thick and thin film; or integrated circuits. Coatings are used in all aspects of printed circuits, from the base substrate materials, through fabrication, to final protection encapsulation of the finished product.

Printed Wiring Boards. Printed wiring boards (PWBs) consist of one or more layers of conductive circuitry adhered to a dielectric (insulating) material or substrate. These may be simple single-sided copper wiring traces on a support, or complex multilayer assemblies having plated-through holes (vias), used in higher performance, higher density packaging (see PACKAGING, SEMICONDUCTORS AND ELECTRONIC MATERIALS).

Substrate Materials. The most common substrate is a copper-clad laminate. Fabrication starts with a base material, usually glass cloth, impregnated with a thermoset resin that is partially cured to give what is called a prepreg. The resin may be epoxy, but for more demanding applications such as the high density multilayer boards, high temperature polyimides are used. The prepreg is then clad with a thin layer of copper foil in a lamination or pressing step to give a rigid substrate suitable for and used in most applications. Demand for space savings has led to increased use of flexible substrate materials. These consist of a thin polyimide dielectric material clad with an adhesive and copper, and can be configured similarly to the rigid PWB (see ADHESIVES).

Resists. Resists are temporary, thin coatings applied to the surface of the copper-clad laminate. After patterning, these films act as masks that are chemically resistant to the cleaning, plating, and etching solutions used to define the circuit traces of the PWB. Both nonphotosensitive and photosensitive types are used.

Screenable Resists. Screenable resists or inks (qv) are applied to the metal-clad substrate through a silk (qv), nylon, or stainless steel screen on which a circuit pattern has been defined. The coating is squeegeed through the stencil onto the substrate, then dried. Depending on whether metal is to be

removed or added, the board is treated with either etching or plating baths or solutions. Afterward, the resist is removed or stripped. Organic soluble resists have almost entirely been replaced by aqueous alkali strippable resists.

Screenable inks have a resin or polymer base and are of three types: organic solvent soluble, aqueous alkali soluble, and permanent. Primarily because of pollution requirements and higher solvent costs, the aqueous types have come into greater use. The permanent types are used as solder masks or for marking the boards. Uv-curable inks are also in use.

Although more direct than photoresist, this method, limited to line widths of 250–380 μm (10–15 mils) because of the limits of screen fabrication, is used primarily for low cost print-and-etch and plated PWBs (5).

Photoresists. For high resolution circuit lines <125 μm (5 mils) photosensitive resists are used. These can either be applied as liquids and dried or laminated as dried films supported on a polyester support. When exposed to light, typically ultraviolet radiation, these coatings change chemically, and their solubility to certain solvents or developers changes. There are two types of photoresist: negative-acting and positive-acting. Negative resists typically consist of a mixture of acrylate monomers, a polymeric binder, and a photoinitiator. Upon uv exposure through a mask, the resist polymerizes and becomes insoluble to the developer. Unexposed areas remain soluble and are washed away. Positive resists function in the opposite way with exposed areas becoming soluble in the developing solvent. As for screenable inks, photoresists are available commercially both as solvent or aqueous developable, as well as permanent. A schematic of negative- and positive-working resists is shown in Figure 3.

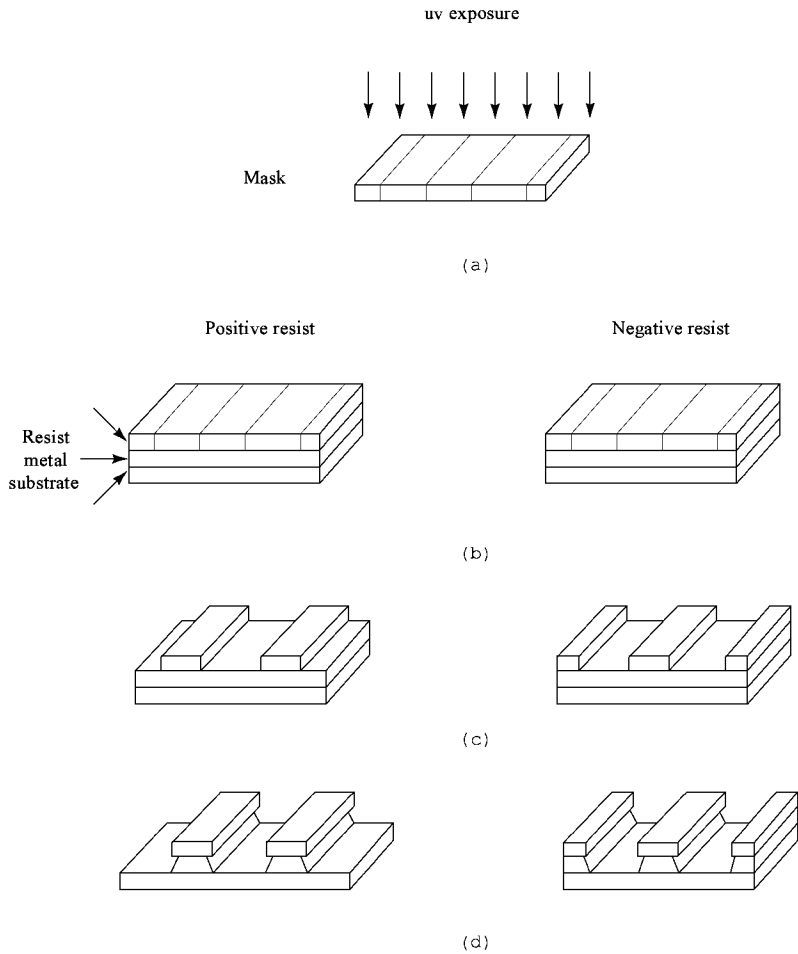


Fig. 3. Lithographic process using positive- and negative-resist systems. The element (a) is (b), exposed to uv radiation; (c), developed; and (d), the metal is etched (6).

Protective Coatings. Solder masks are resists designed to define or mask areas of a finished printed circuit board where components are to be attached by selective deposition of molten solder (see **SOLDERS AND BRAZING ALLOYS**). In contrast to etching or plating resists, the cured solder mask film remains permanently. These coatings minimize bridging between close conductor lines and pads during soldering, provide protection from fluxing and other chemical attack during fabrication, and act as environmental and physical barriers for the life of the board. This is a demanding application having a number of rigidly specified requirements such as abrasion resistance, flexibility; adhesion to copper, tin-lead, nickel, and gold; thermal and flame resistance; machinability; low water absorption; and chemical and solvent resistance.

Earlier material systems were screened epoxies, but both laminatable dry-film and liquid uv photoimageable coatings are finding wide acceptance. These radiation-cured coatings offer high resolution and registration accuracy, and can be either epoxies or epoxy acrylates.

After all the components have been mounted, a conformal protective coating may be applied to the printed circuit board not only to provide electrical insulation, but also to help maintain performance and extend service life in harsh operating environments. Moisture is the main culprit, causing corrosion, degradation of insulating properties, or electrical shorts. Other benefits are protection from contaminants such as spray, salt, dust, grease, fuel, corrosive vapors, fungi, and, to some extent, mechanical vibration. Many electronic component assemblies are inexpensive and designed to be nonrepairable and throwaway. For these, if a protective coating is used at all, it is an inexpensive varnish or polyester. For more expensive modules, specialty coatings are used to ensure long-term reliability. These are typically acrylic, epoxy, polyurethane, silicone, or polyxylylene in composition and can be applied by numerous methods, eg, dip-coating, spraying, fluidized-bed coating, casting, or vacuum deposition. Individual components themselves may also be protected using conformal coatings. For expensive assemblies, the coating must be removable to allow solder rework or replacement of defective components. This requirement places further demands on the coatings.

Hybrid Microelectronic Coatings. Hybrid circuits are fabricated using two complementary technologies: thick- and thin-film printed circuits. Typically, simple thick- and thin-film hybrid circuitry and components packages, including integrated circuits (ICs), are integrated assembled into a complex integrated unit on a ceramic substrate. The function of the hybrids is to connect and modify the component packages through use of various conductors, dielectrics, resistors, inductors, and capacitors. Because these can function at high temperatures and be hermetically sealed for some applications, they are typically used in demanding applications such as for the military, in space, computers, the automotive industry, and telecommunications.

Thick-film technology is an additive process and involves screen printing of functional inks, also called pastes or simply compositions. The paste formulation consists of a high (20,000–250,000 mPas(cP)) viscosity thixotropic dispersion of fine particles of an electrically functional phase and a glass and or ceramic frit binder. Also included for screen printing are a polymeric binder, a solvent or vehicle, and modifiers such as wetting and flow-control aids. After air drying to remove most of the solvent, the parts are fired at high (850–1000°C) temperature. During this process, the polymeric binder and

other small molecule species decompose and volatilize. The glass frit fuses, wetting the surface of the functional phase, providing adhesion and sealing of the composite to the substrate. Because of the screen printing process, resolution is modest. Fired film thicknesses, which range from 10 to 50 μm (0.4 to 2.0 mils), are large compared to thin-film microelectronics. Some photosensitive pastes are also in use.

Polymer thick films also perform conductor, resistor, and dielectric functions, but here the polymeric resins remain an integral part after curing. Owing to the relatively low (120–165°C) processing temperatures, both plastic and ceramic substrates can be used, leading to overall low costs in materials and fabrication. A common conductive composition for flexible membrane switches in touch keyboards uses fine silver particles in a thermoplastic or thermoset polymeric binder.

Thin-film technology offers the highest circuit feature resolution, down to micrometer level geometries or even below, and film thicknesses that range from 0.01 to several micrometers (see THIN FILMS). Although used in hybrid manufacturing, these processes are more typically found in the fabrication or manufacture of microelectronic, ie, integrated circuit devices. Coating is primarily done by either of two vacuum techniques: evaporation or sputtering. Microelectronic photoresist application by spin coating and photodefinition is followed by etching, a subtractive process, to yield the circuit and device features. This contrasts with thick-film processes, which are additive. However, for many magnetic and optical applications, the thin film may be deposited only where desired through a stencil as in screen printing.

Device materials again may be conductive, semiconductive, dielectric, or resistive. Conductors are typically gold or aluminum, and resistors, silicon monoxide or silicon nitride. Tantalum nitride and nickel chromium are common resistor materials.

Vacuum evaporation consists of heating a reservoir of the source material to its boiling or sublimation point in a high vacuum. Material vapor travels a distance and condenses on a relatively cold substrate, forming a very thin uniform film. In sputtering, energetic ions are accelerated into a target material at relatively high pressures, and particles are kinetically driven from the surface. Several newer techniques are plasma polymerization and chemical vapor deposition (see PLASMA TECHNOLOGY; VACUUM TECHNOLOGY) (7,8).

Numerous protective conformal coatings are also used in the hybrid circuit industry. Most of the criteria that pertain to use in printed wiring also apply here (4).

Integrated Circuit Coatings. The integrated circuit (IC) is the basic building block in microelectronics. It is a semiconductor device having both passive and active components built both into and onto a silicon wafer. High reliability has led to numerous applications in aerospace, and for industrial and consumer equipment. Precisely formed multiple layers are deposited on top of one another and interconnected to form a three-dimensional circuit network. Circuit geometries are generally measured in the micrometer range. Circuit element thicknesses vary from a few hundredths to a couple of micrometers. A large integrated circuit chip often contains in excess of a million devices. Individual chips are linked together through packages on printed wiring boards or hybrid substrates to communicate with one another and the outside world.

Many of the fabrication processes for integrated circuits are similar or conceptually related to those used in the manufacture of printed wiring boards. However, because of the extremely fine device features, fabrication must be carried out in clean rooms having strictly controlled environments. Particulate and chemical contamination are minimized, and temperature, humidity, and even vibration are carefully controlled.

Resists used to define circuit patterns are radiation-sensitive and may be either positive- or negative-working. As a result of the fine lines, there has been movement away from optical lithography and into the mid- or deep-uv regions. Developmental work has also been focused on electron beam, x-ray, and ion-beam exposure devices and resists (9,10).

Polymides, both photodefinable and nonphotodefinable, are coming into increased use. Applications include planarizing interlayer dielectrics on integrated circuits and for interconnects, passivation layers, thermal and mechanical stress buffers in packaging, alpha particle barriers on memory devices, and ion implantation (qv) and dry etching masks.

A number of polymeric coatings, primarily polyimides, epoxies, and silicones, are used as adhesives and in other aspects of packaging, as well as for final encapsulation and protection of the integrated circuit.

Coating Application Methods

There are a wide variety of coating applications processes in use. The majority of these techniques are similar to those used in other coatings industries, and the same basic operating principles apply to these uses as to coating a photographic film or a coil of metal for a refrigerator.

There are, however, several differentiating characteristics of electronic coating processes as compared to general use. Typically, more discrete objects are coated as opposed to continuous webs, although wire coating, optical fiber, magnetic tape, metallized polyester, and optical fibers are continuous webs coated at very high volumes. A much wider range of coating weights and coating coverage is needed. The coatings and line width in an integrated circuit are very thin and low in coating weight. The protective coatings are relatively heavy, thus a wider variety of coating methods are in ongoing use in the electronics manufacturing industry than in most other industries. Spray, dip, brush, roller, fluidized-bed, spin, electrocoat, and vacuum deposition are all in routine use. Table 3 summarizes these methods for electronics coating applications.

Table 3. Electronics Coatings Application Methods^a

Coating method	Advantages	Limitations	Typical uses
spray	fast, adaptable to varied shapes, low cost	difficult to obtain complete coating thickness, not uniform, poor reproducibility, high coating solution loss, solvent loss	motor frames and housings, electronic enclosures, circuit boards, modules, protective decorative coatings
electrostatic spray	efficient coverage complex shapes, automated, less solvent loss	requires specially formulated solutions, high cost	heat dissipaters, decorative and protective coatings, enclosures
dip	thorough coverage of complex parts, inexpensive	viscosity stability affects coverage	protective coatings on small- and medium-sized parts
roll	high speed continuous process sheets, excellent thickness control	uniformity, environmental concerns	protective and decorative coatings, application of liquid photoresist to boards
fluidized-bed	thick coatings in one pass, uniform coating, no solvent, environmentally friendly	requires continuous web, high cost	motor stators, insulation on castings, metal substrate for heat sinks, circuit boards
screen stencil	deposits coating in select area through mask, good thickness control	high temperature needed to fuse coating, limits substrate material	circuit boards, masking for etchants, spot insulation between circuitry layers
spin	good thickness control for low coverage, reproducible	requires flat surface, discrete substrate, not continuous	photoresists for semiconductor devices and thin-film circuits, polymeric dielectrics, magnetic disks
electrocoat	good control of thickness and uniformity, can coat wet parts, good complex shapes	limited number of coating formulations, must be specially formulated, adhesion	protective and decorative coatings, complex castings
vacuum deposition	ultrathin, pinhole-free coatings, selective deposition through masks	thermal instability of plastics, needs vacuum control	optical coatings, thin-film IC layers

^a Ref. 10.

In addition, cleanliness is far more important at all stages of the electronics coating process than generally necessary for nonelectronics coatings. There is also a wider range of coating compositions used, and because most of the processes are fluid coating application processes, many diverse solvents are in use. This contrasts with the paper and photographic industries where water is almost always the exclusive solvent used. The wide use of high volatility solvents in a coating application is of concern, however, because of the increasing environmental needs.

BIBLIOGRAPHY

1. *Electronics*, 26 (Jan. 1992).
2. D. J. Broer, *Adv. Org. Coat. Sci. Technol. Ser.* **11**, 219–228 (1989).
3. J. E. Sturge, V. Walworth, and A. Shepp, eds., *Imaging Process and Materials*, 8th ed., Van Nostrand Rheinhold, New York, 1989, p. 252.
4. J. J. Locari and L. A. Hughes, *Handbook of Polymer Coatings for Electronics*, 2nd ed., Noyes Publications, Park Ridge, N.J., 1990, p. 224.
5. C. F. Coombs, Jr., *Printed Circuit Handbook*, 2nd ed., McGraw-Hill Book Co., New York, 1979, pp 6-9–6-31.
6. Ref. 3, p. 569.
7. K. Bunshah and co-workers, *Deposition Technologies for Films and Coatings*, Noyes Publications, Park Ridge, N.J., 1982.
8. H. K. Pulker, *Thin Films Science and Technology*, Elsevier, Amsterdam, 1985.
9. D. S. Soane and Z. Martynenko, *Pofymers in Microelectronics: Fundamentals and Applications*, Elsevier, Amsterdam, 1989.
10. M. S. Htoo, *Microelectronic Polymers*, Marcel Dekker, Inc., New York, 1989.

Edward Cohen
Howard E. SimmonsIII
Du Pont Central Science & Engineering



ELECTRONIC SUBSTRATES.

See INTEGRAL CIRCUITS; SILICON.

ELECTRON SPIN RESONANCE.

See MAGNETIC SPIN RESONANCE.

ELECTROPHORESIS.

See ELECTROSEPARATIONS, ELECTROPHORESIS.

ELECTROPHORETIC DEPOSITION.

See COATING PROCESSES.

ELECTROPHOTOGRAPHY

The first xerographic copy was made in 1938 (1). Triboelectricity, the charging of solids by friction, was first reported by the Greeks. Not until the early twentieth century were there serious efforts to develop what became known as electrography for reproducing images. The production of such images was based on the ability of electrostatically charged insulators to attract triboelectrically charged powders (2). This required the scanning of a document, line by line, using a light-sensitive scanner to produce an analogue voltage that was then used to modulate the output from an ion source. Copies could thus be produced by the displacement of a document-size, insulating layer beneath a stationary ion source and then rendering the resultant charged image visible (developed) by dusting with a triboelectrically charged powder.

The combination of electrostatic charging and development using the phenomenon of photoconductivity, first discovered in selenium crystals in the 1870s, led to the invention of electrophotography (3,4). Photoconductivity, the enhancement of the electrical conductivity of a material by illumination with light, allows the image of an entire document to be copied by simply projecting its optical image onto a page-size photoconductive layer, known as the photoreceptor, uniformly charged with ions. Reflected light from the document then produces selective photodischarge in the photoreceptor proportional to the incident light intensity. The resultant image, consisting of the remaining surface charge, replicates the information content of the document. This latent electrostatic image can then be developed by utilizing its electrostatic attraction for charged powder. The highly photosensitive amorphous form of selenium [7782-49-2] was discovered in the mid-1940s. Other improvements involving ion charging processes, electrostatic transfer, and dry-ink or toner materials followed. Commercialization became feasible and the Greek words for dry, xeros, and writing or drawing, graphein, were combined to give the name xerography and XeroX as a trade name.

The XeroX Copier Machine Model A was announced in 1949, and involved complicated manual operation. Copies of acceptable quality were operator dependent. The Copyflo printer, introduced in 1955, was the first automated xerographic machine and enabled the production of copies on a continuous web of ordinary paper. Early electrophotographic products used paper coated with dye-sensitized zinc oxide Electrofax which had met market resistance in terms of aesthetics and cost, so that in 1958 the total market was only about \$100 million (1–3,5).

The Xerox 914 copier, first shown in September 1959, was a fully automatic machine requiring no special skills to operate. It produced seven copies per minute on plain paper using a reusable photoreceptor. The image, composed of a carbon-black impregnated polymer toner, was impervious to degradation by light or chemicals and thus compatible with archival storage. In contrast to market predictions of 3,000, over 200,000 were sold over the life of the product. By 1992 the total copier and printing business worldwide for all companies had grown to over \$50 billion and the annual total number of copies was numbered in trillions. Developments in materials and process, including silicon integrated circuit technology (see INTEGRATED CIRCUITS), software, digital transmission networks, and light sources, such as lasers (qv) and light-emitting diodes (LED) (see LIGHT GENERATION), have extended the capability of xerography beyond making copies to creating new documents based on the synthesis and manipulation of electronic images.

The features of the integrated xerographic process are indicated in Figure 1 (3). The heart of a xerographic machine is the photoreceptor which consists of a thin film, 20 to 50 μm thick, coated onto a grounded aluminum drum or metallized belt (see THIN FILMS). The film material must have a high ($>10^{12} \text{ } \Omega\text{cm}$) resistivity in the dark and yet be photosensitive on exposure to visible light. The photoreceptor drum or belt is mechanically rotated with respect to the other necessary subsystems. The first of these is a corona charging device, such as a corotron, which produces either positive or negatively charged air ions depending on the polarity of a high voltage applied to a wire. These ions are accelerated by the applied voltage and deposited on the surface of the photoreceptor, which thus charges as a parallel plate capacitor. The resultant voltage across the film is proportional to the number of ions deposited per unit surface area and to the inverse of the material's dielectric constant. Typically, voltages of a few hundred volts are used to provide the necessary development fields. This surface potential, divided by the film thickness, determines the applied electric field within the photoreceptor. The charged area of the photoreceptor then rotates, with minimal loss in the initial voltage because of its high resistivity in the dark, into the exposure position.

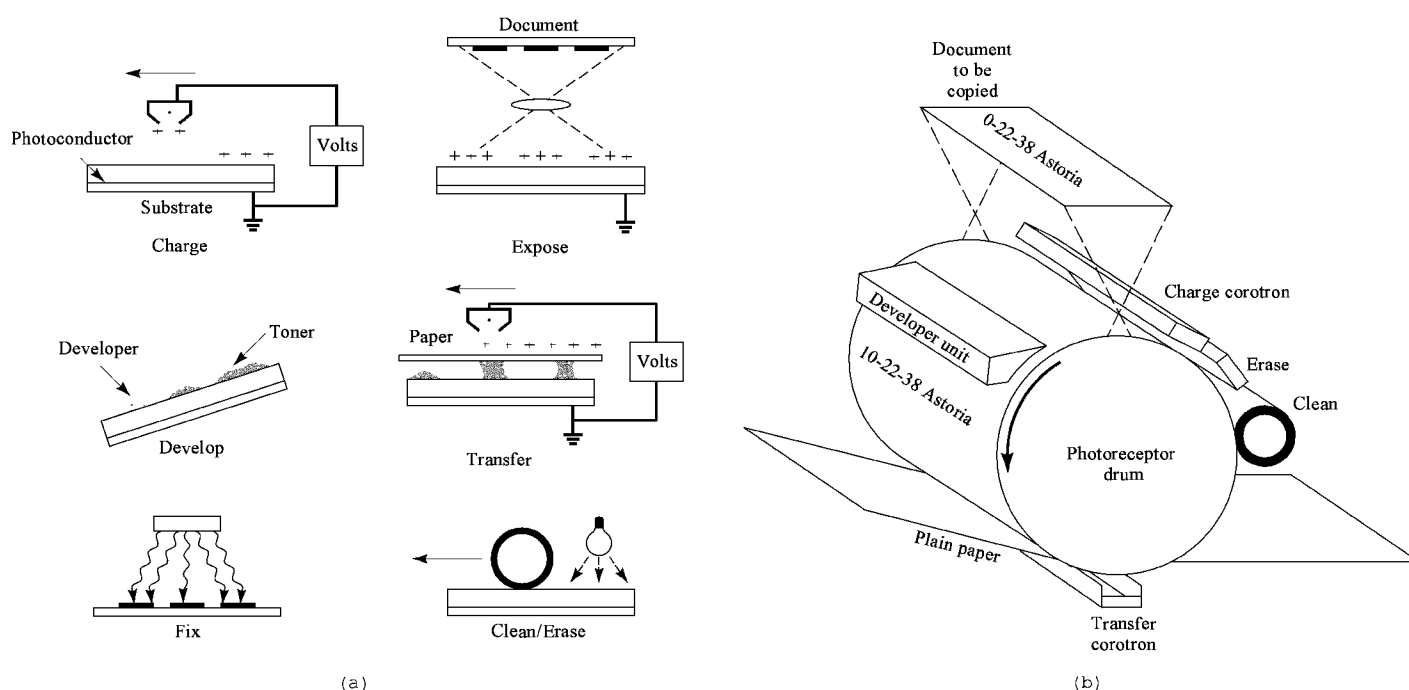


Fig. 1. (a) Steps of the xerographic process; (b) schematic arrangements of hardware.

In the exposure position an optical image, produced by scanning the document to be copied with a slit light source, is focused by lenses onto the photoreceptor surface. Where light reflected from white areas of the document strikes the photoreceptor, photogeneration of free electrons and holes occurs within the absorption depth of the light, which is typically much less than the film thickness. The dark characters on the original document reflect less light and proportionally fewer photocarriers are created. Assuming the initial voltage is positive, photogenerated holes move across the photoreceptor film and electrons move the shorter distance back to the top surface under the influence of the electric field within the photoreceptor. This causes the initial voltage to selectively discharge to a degree depending on the amount of light locally absorbed by the photoreceptor. In laser printers or digital copiers, the optical image is directly written on the photoreceptor surface by a scanning laser or LED array where the light output is controlled in an on-off fashion using digital electronic signals.

Leaving the exposure zone, the photoreceptor now carries an invisible distribution of surface charge or voltage which replicates the document image. This latent electrostatic image is then developed by cascading charged dry ink or toner, consisting of polymer particles containing carbon black or colored pigment, over the photoreceptor surface. The toner charge which is produced by controlled triboelectric charging is opposite in sign, negative in this example, to the original charge deposited on the photoreceptor. The negatively charged toner particles are attracted by mutual coulomb electrostatic force to the remaining positive surface charges of the latent electrostatic image, and become attached to the photoreceptor surface.

After leaving the development zone, the original document charge image is now a visible toner image. This is next transferred from the photoreceptor to plain paper by using a transfer corotron. This charges the paper with ions of the same polarity as that used to charge the photoreceptor, and toner now moves from the photoreceptor onto the paper. Because only a fraction of the toner is transferred in this step, the remainder must be removed from the photoreceptor surface, for example, by mechanical brushing. Finally, any residual latent-image surface charge must be discharged. This is done by uniformly exposing the entire photoreceptor to light from an intense erase lamp. After the transfer step, the paper moves to a fixing or fusing station where the toner image is softened using either solvent or, more typically, heat. In this way the toner adheres to the paper, the image is fixed, and the copy is ready. At this point the process cycle is completed and can automatically restart. In modern high speed xerographic products, these steps occur automatically to produce over 100 copies or prints per minute.

Photoreceptor Materials

The photoreceptor employs photoconductive materials and basically serves the same function as silver halide film in conventional photography (qv). The photoreceptor is designed to transform an optically input image into an electrostatic latent image on the photoconductor surface. The photoconductor remains charged if unilluminated, or even when exposed to unabsorbed wavelengths of light, discharging in the regions where light is absorbed.

The most desirable characteristics of photoreceptors for use in electrophotography are the following:

- (1) The ability to accept and hold the electrostatic charge in the darkness. The photoconductive layer should support a surface charge density of approximately $0.5-2 \times 10^{-7} \text{ C cm}^{-2}$. The charge also has to be uniformly distributed along the surface, otherwise nonuniformities can print out as spot defects. The applied surface potential should be retained on the photoreceptor until the time when the latent electrostatic image is developed and transferred to paper or, if needed, to an intermediate belt or drum. In other words, the "dark decay" or conductivity in the dark must be very low. The photoconductor materials must be insulators in the dark.
- (2) The photoreceptor must lose most of its surface charge upon exposure to the white light reflected from the original or upon exposure to an attenuated laser beam. The residual potential after each exposure must be low, not greater than a very small fraction of the original surface potential. The amount of available light energy is small, typically less than about $1 \times 10^{-6} \text{ J cm}^{-2}$ (10 ergs cm^{-2}) for fast duplicators and printers and somewhat more for slower copiers and printers. Thus the device should absorb as much light as possible, and the number of charges generated and transported without a loss per incident photon must be as high as possible. This means high absorptivity of the material designed to generate charges, high charge-carrier generation efficiency, no loss of charge carriers by recombination or from traps when the carriers are transported across interfaces, and essentially trap-free transport.
- (3) The photogenerated charge carriers should exit the device in a time shorter than the time between exposure and the development of the latent image. In practical terms, the charge-carrier mobilities μ (velocities v per unit electric field E), should be of the order of $10^{-7} - 10^{-6} \text{ cm}^2 (\text{Vs})$ for slow copiers and printers, or $10^{-6} - 10^{-5} \text{ cm}^2 (\text{Vs})$ or higher for high volume copiers or printers. It is also important that essentially all charges exit the device in each xerographic cycle, ie, no charges should be accumulated by deep trapping. Deep traps are operationally defined as traps from which the carriers are released in times longer than the time of the experiment, which in this case is the xerographic cycle. This places stringent requirements on the purity of all participating materials, including the solvents used for the deposition, and on the chemical stability of all the photoreceptor components; the presence of as few as 10^{13} trapping defect molecules per cubic centimeter can noticeably affect the residual potential on the photoreceptor.
- (4) The photoreceptor must be stable in performance. In order to guarantee that the quality of copies stays the same after hundreds of thousands of produced copies, the photoreceptor should maintain the same charge acceptance. That is, the potential across the photoreceptor should be the same

after each charging providing the charging device, the corotron, produces the same number of ions. The rate of the surface charge decay should not increase with the xerographic cycling. In other words, the photoconductor conductivity in the darkness should not vary with time. Also, the number of charges produced and transported (the sensitivity) per incident photon should remain the same. This means that the device should not undergo any degradation process which would result in an increase of the surface or bulk conductivity, in the development of charge traps, in a change in sensitivity, in a change of the blocking nature of the metal–photoconductor interface, or in a significant change of fatiguing. Fatiguing is a temporary increase in dark conductivity after several successive xerographic exposure cycles, which disappears after a brief rest in the darkness. Charge trapping could be caused by impurities, either inherent or formed during the xerographic process.

- (5) The photoreceptor should be mechanically robust. This term was coined to reflect the combined mechanical durability, high scratch- and abrasion-(wear) resistance, high flexibility, resistance to cracking and crazing if the photoreceptor is used as a flexible belt or scroll, smoothness, which is believed to enable efficient cleaning of residual toner particles, as well as a chemical resistance to corotron effluents containing ozone, oxides of nitrogen, and other corrosive elements, compounds, or ions.
- (6) Lastly, the photoreceptors must be inexpensive and easy to fabricate into defect-free, large-area thin films with uniform thickness of all layers.

Photoconductors. Only the most recently developed organic photoconductors (OPC) approach optimum behavior (see PHOTOCONDUCTIVE POLYMERS).

The early xerographic photoreceptors were single-layer devices, 20–100 μm thick, comprised of a photoconductive material deposited on a conductive ground, a drum or a plate, as illustrated in Figure 2a. Because most photoconductive materials have high absorption coefficients, the incident photons are absorbed and, consequently, most of the charge-carrier pairs, holes and electrons, are produced in a very thin surface region. The sign of charge deposited on the photoconductor depends on which sign carrier has the best chance of moving through the bulk of the material without trapping, and which is faster. Thus amorphous selenium-based photoreceptors are charged positively because selenium transports holes better than electrons. The photogenerated electrons, the role of which is to neutralize the surface charge, may have a limited range but do not have to travel a large distance.

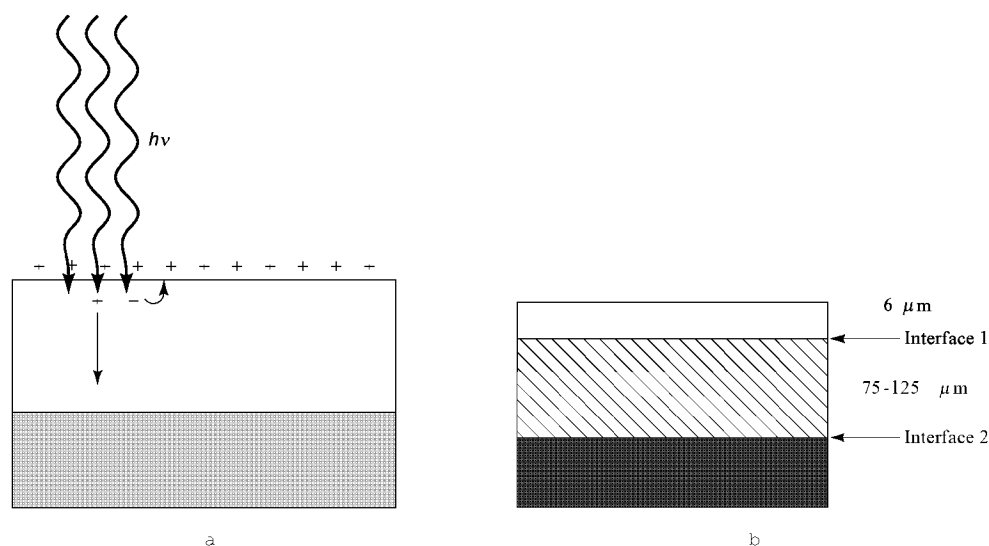


Fig. 2. Schematics of (a) single-layer photoreceptor, where the + signs represent the corona-deposited charge, □ the photoconductor, and ■ the conductive substrate; and (b), the $\text{CdS}_x\text{Se}_{1-x}$ (Katsuragawa) photoreceptor, where □ represents the insulating layer, ▨ the $\text{CdS}_x\text{Se}_{1-x}$, and ■ the conductive base.

Anthracene and sulfur were used as photoconductors to produce the first xerographic images in the late 1930s (1,4,6–9). These elements do not easily form good thin films nor absorb visible light strongly enough.

Amorphous (vitreous) selenium, vacuum-deposited on an aluminum substrate such as a drum or a plate, was the first photoconductor commercially used in xerography (6). It is highly photosensitive, but only to blue light (2). Its light absorption falls off rather rapidly above 550 nm. Because of the lack of photoresponse in the red or near infrared regions, selenium photoreceptors cannot be used in laser printers having He–Ne lasers (632.8 nm), or solid-state lasers (680–830 nm).

The useful spectral range of selenium can, however, be extended. For example, the addition of substantial amounts of tellurium shifts the absorption tail to the red or eventually to the near infrared (10). Addition of 7.5% of Te extends the spectral response almost throughout the visible region. Further increase of the Te content, however, may lead to compositional instabilities and phase separation, particularly when the alloy is exposed to corona or to an abrasive environment. The localized change in composition results in higher dark conductivity and the localized loss of the surface potential, which can lead to image degradation. One way of minimizing such compositional instabilities is to build a multilayer structure comprising a thick layer of selenium in conjunction with a thin Se–Te layer (11,12). Se–Te alloys in layered configurations have been used in a number of copiers.

Another disadvantage of selenium is that it crystallizes readily when heated above about 55°C. The crystallized form of selenium is highly conductive and therefore when the photoreceptor is accidentally exposed to higher temperatures, it does not accept charge. Crystallization can be retarded by the addition of small amounts of arsenic (several wt % based on selenium) (13). The addition of larger amounts of As also extends the spectral response into the red region (14). Selenium alloy photoreceptors having about 40 wt % As (nearly stoichiometric As_2Se_3) are highly sensitive throughout the visible region of light and have been used in high speed laser printers in conjunction with the He–Ne lasers. Only holes are mobile in this photoconductor; therefore, these devices must be charged positively. The hole mobility μ in what is essentially arsenic triselenide [1303-36-2], As_2Se_3 , is high, and therefore the material is well suited for high speed duplicators or laser printers. For example, at a typical xerographic electric field of 10^5 V/cm at 293 K, the hole mobility μ is $7.5 \times 10^{-5} \text{ cm}^2/(\text{Vs})$ (15). These photoreceptors are also characterized by unusually high abrasion resistance and environmental stability. The principal drawback, however, is the toxicity of arsenic.

Zinc oxide was also quite successful in xerographic photoreceptors, particularly in Electrofax (5). In this process the image is developed and fixed by fusing directly into a ZnO-sensitized paper. Crystalline ZnO is used as a fine dispersion in a binder host polymer such as poly(vinyl acetate), poly(vinyl butyral), polyacrylics, and the like, in about 5 to 1 weight ratios, close to 1:1 by volume (17). These photoreceptors are usually charged negatively, because the layers do not retain enough positive charge. The ZnO devices have a peak photoresponse between 370 and 390 nm (18). To shift the absorption to the visible range, sensitizers such as Rose Bengal [11121-48-5], $\text{C}_{20}\text{H}_4\text{Cl}_4\text{I}_4\text{O}_5$; fluorescein [2321-07-5], $\text{C}_{20}\text{H}_{12}\text{O}_5$; Brilliant Green [633-03-4], $\text{C}_{27}\text{H}_{34}\text{N}_2\text{O}_4\text{S}$; acridine orange [65-61-2], $\text{C}_{17}\text{H}_{19}\text{N}_3 \cdot \text{HCl}$; rhodamine B [81-88-9], $\text{C}_{28}\text{H}_{31}\text{ClN}_2\text{O}_5$; bromophenol blue [115-39-9], $\text{C}_{19}\text{H}_{10}\text{Br}_4\text{O}_5\text{S}$; etc, are added (18,19). Some of these devices have a respectable quantum efficiency of charge generation and collection, approaching 0.4 (20). The nature of the polymeric binder has a large effect on the device performance (21), and so does the quality and source of the dye (22). Sensitivity to the environment and fabrication methods results in some irreproducibilities and batch-to-batch variances. However, the main advantage of the ZnO-based photoreceptor paper is its very low cost.

Other photoconductive pigments (qv) which found commercial application are cadmium sulfide [1306-23-6], CdS, or the alloy with Se, CdS_xSe_{1-x}. This latter pigment was used in several electrophotographic processes (23,24) in which the photoreceptors have one feature in common: each contains the powder pigment, ie, the polymer binder matrix, sandwiched between the conductive ground and an insulating overcoat layer (see Fig. 2a) (25). These processes differ from the conventional electrophotographic processes using selenium or ZnO. In one of the latter imaging methods (23), the insulating surface is first charged positively, inducing a negative charge in the photoconductive layer at interface 1 of Figure 2b. In the second step, the device is charged negatively by corona while it is exposed simultaneously to the projected image. In this step, electron-hole pairs are generated in the photoconductive layer, near interface 1. Because the field is now reversing, holes are trapped in that region, and electrons migrate to the ground. The deposited surface charge in the dark portions of the image is thus reduced to very low levels, whereas in the illuminated portions the negative charge stays on the surface. In the last step, the device is uniformly illuminated. Electrons, which are trapped in the dark portion of the photoconductor layer, are neutralized by photogenerated holes, and newly generated electrons migrate to the ground. The latent electrostatic image on the surface is then developed as in conventional xerography.

The advantage of this process over other xerographic processes using conventional photoconductors is that the most sensitive part of the photoconductor is protected from the corrosive environment. This helps to increase the useful life of the device. The process itself, however, is rather complicated.

A highly sensitive CdS pigment was also developed for a conventional xerographic application. Optimum xerographic properties were obtained using copper doping at Cu concentrations of 200–300 ppm (26).

Amorphous Silicon. Significant effort has been expended in the development of an amorphous silicon photoreceptor (see SILICON AND SILICON ALLOYS). For xerographic application, the α-Si:H must be insulating, the permissible conductivity σ no greater than about 10⁻¹⁰ (Ωcm)⁻¹. Pure amorphous silicon without hydrogen is too conductive because of the presence of large numbers of dangling bonds. Hydrogen, incorporated into the structure, when prepared by glow discharge of silane gas, tends to saturate the dangling bonds, but even with 10–35% hydrogen atoms, α-Si:H is still too conductive for xerography and must be slightly p-doped with boron, up to ca 10-ppm levels (27,28). The hole mobility in lightly boron-doped α-Si:H is ca 10⁻³ cm² (Vs)⁻¹, which is more than sufficient for even the fastest duplicators and printers (see Fig. 3). Boron doping is accomplished by adding diborane [19287-45-7], B₂H₆, to the reacting gas (see BORON COMPOUNDS). The discharge is excited in a capacitively or inductively coupled reactor using d-c, r-f, or microwave power sources.

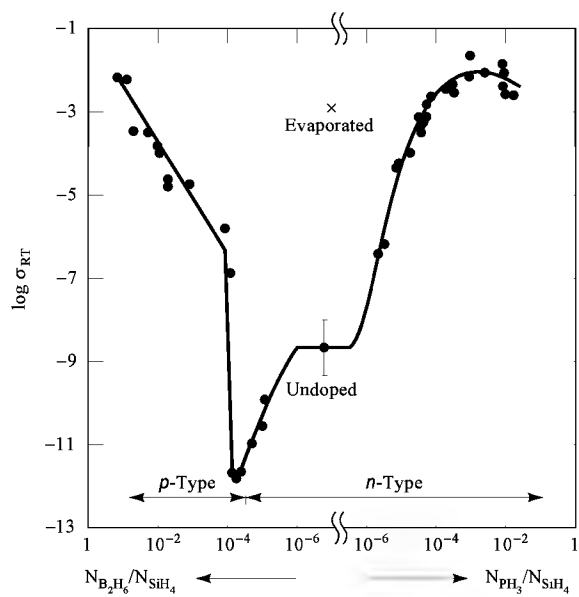


Fig. 3. Logarithm of room temperature electrical conductivity of α-Si:H as a function of doping with diborane, B₂H₆, and phosphine [7803-51-2], PH₃, where N_{B₂H₆}/N_{SiH₄} is the ratio of the number of diborane to silane molecules; N_{PH₃}/N_{SiH₄} is the ratio of phosphine to silane molecules. Both ratios refer to the gas mixtures used for sample preparation.

The main advantage of α-Si:H is the long mechanical life and sensitivity in the whole visible spectrum. However, the primary failure mode of the unprotected α-Si:H photoreceptors is not the reduction of thickness by wear and abrasion, but a catastrophic failure attributable to microscopic scratches caused by foreign objects, dust, or paper, or to corona damage. Gap states can be introduced easily to α-Si:H either as a result of structural irregularities of the damaged surface or by spontaneous oxidation. These new gap states at the surface are responsible for generation of dark charge carriers, and the microscopic damage thus shows in the xerographic copy. It is necessary to overcoat α-Si:H using a thin hard shell of nonstoichiometric silicon carbide [409-21-2], SiC, or silicon nitride [12033-89-5], Si₃N₄. Also, because the α-Si:H photoreceptors are charged positively, spontaneous dark injection of electrons from the conductive substrate must be prevented because such dark-injected electrons would neutralize the surface charge before the latent image could be developed. Electron injection is blocked by a thin layer of α-Si:H heavily doped with boron. As shown in Figure 4 about 1000 ppm boron is inserted between the electrode and the bulk of α-Si:H. Boron doping reduces the electron range.

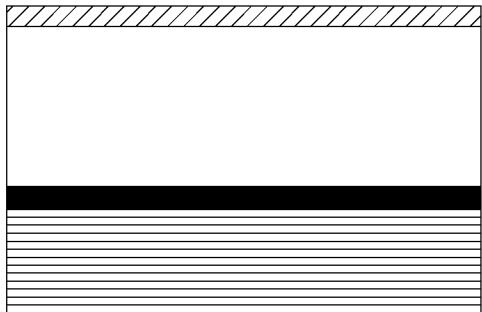


Fig. 4. Structure of an α-Si:H photoreceptor, where ▨ represents SiC_x or SiN_x; □, α-Si:H containing ca 10 ppm B; ■, α-Si:H with ca 1000 ppm B; and ▨, the conductive substrate.

α-Si:H has sufficient sensitivity to He–Ne lasers (632.8 nm) but lacks the sensitivity to solid-state lasers (780–830 nm) (see LASERS). The spectral

response of α -Si:H can be extended to the near ir range by adding germanium. This is accomplished by codeposition of silane in the presence of germane [7782-65-2], GeH_4 . Doping with Ge, however, causes an undesirable increase of dark conductivity. This effect is partially mitigated by forming a separate thin layer of α -Si:Ge:H in conjunction with Ge-free α -Si:H, as in the case of alloying selenium, Se, and tellurium, Te.

Organic Photoconductor-Based Photoreceptors. Organic photoconductor-based photoreceptors (OPCs) appeared in xerographic copiers in the early 1970s. Most modern xerographic copiers, duplicators, and printers use OPCs. The first all-organic photoreceptor was a single-layer device consisting of an electron donor polymer, poly(*N*-vinylcarbazole) [25067-59-8] (PVK), mixed with an electron acceptor, 2,4,7-trinitro-9-fluorenone [129-79-3] (TNF), $\text{C}_{13}\text{H}_5\text{N}_3\text{O}_7$, (29). This charge-transfer complex absorbs light and photogenerates charge carriers throughout the visible region. The photoconductor is bipolar, which means that it is capable of transporting both holes and electrons and, in principle, can be charged both positively and negatively. In this polymeric charge-transfer complex, holes are transported by a hopping mechanism via the carbazole groups, and electrons migrate via both complexed and uncomplexed TNF molecules (2). The absorptivity varies with the wavelength between $5 \times 10^{-3} \text{ cm}^{-1}$ and 10^2 cm^{-1} . At 550 nm, for example, the photogeneration efficiency of PVK–TNF at the optimum ratio is 0.1 for an electric field of about $5 \times 10^5 \text{ V cm}^{-1}$ (30). The PVK–TNF photoreceptor, considered to be fairly sensitive, was very successful until it was discovered that TNF is a potent carcinogen. Copiers that used the material have been withdrawn from the market.

Another organic photoconductor which found application in copiers and duplicators is a single-layer, two-phase device comprising a hole transporting molecule, bis[4-(diethylamino)-2-methylphenyl]phenylmethane (TPM) (Fig. 5) molecularly dispersed in a host bisphenol A polycarbonate binder. Coincidentally, polycarbonate also forms a unique cocrystalline complex with a thiapyrylium dye (see Fig. 6) (31). The charge carriers are photogenerated in the bisphenol A–dye complex which also supports transport of photogenerated electrons. The charge collection efficiency, near 0.6 upon exposure to white light at an electric field of 10^5 V cm^{-1} , is quite high in this photoreceptor. Because both holes and electrons are mobile in this system, in principle, charging of either polarity could have been used; however, in practical applications the device is charged negatively. The only shortcoming of this photoreceptor is the uneven absorption characteristics over the visible spectral range, which leaves the device almost blind at certain wavelengths.

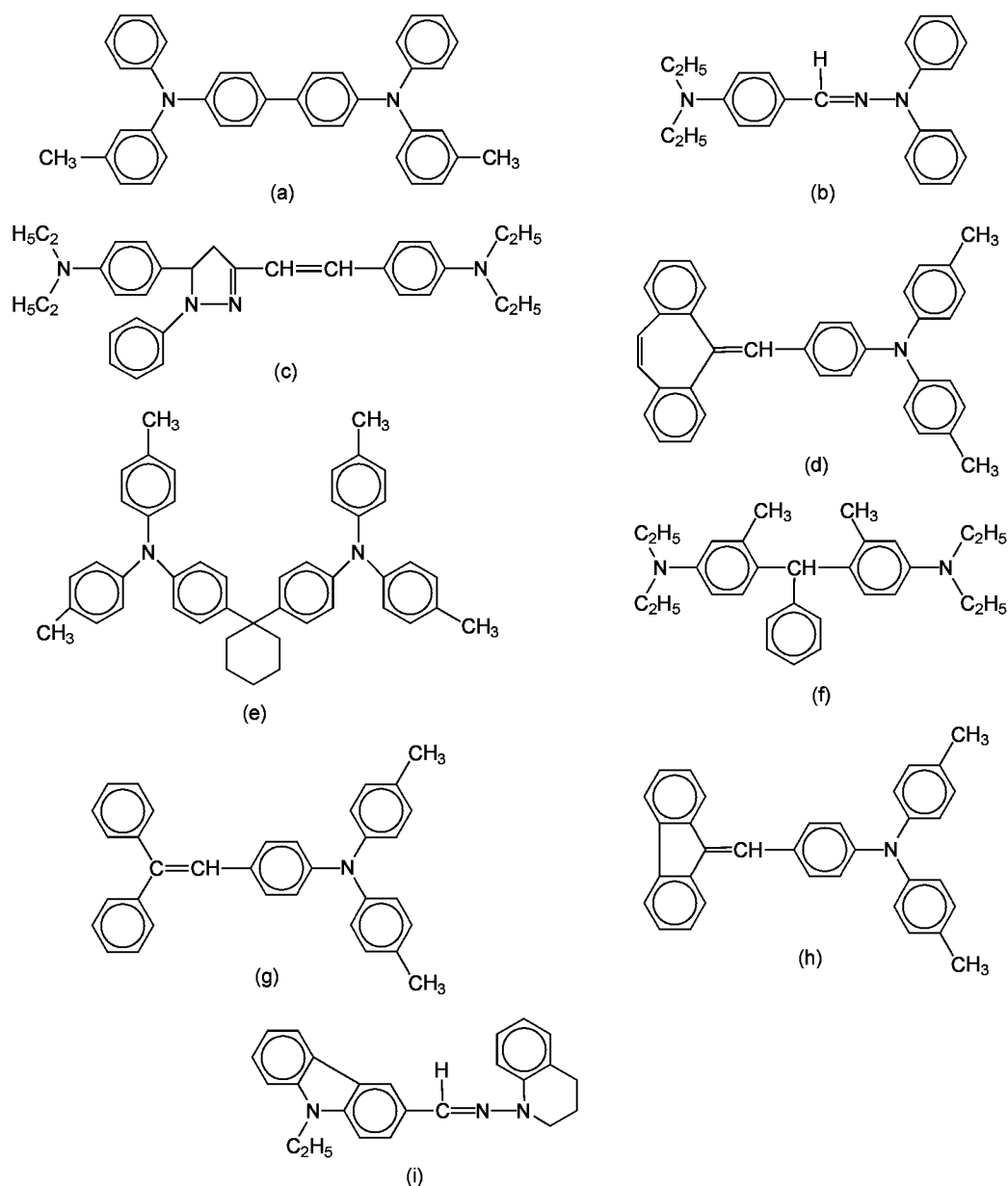


Fig. 5. Charge-transport molecules: (a) *N,N'*-bis(3-methylphenyl)-*N,N'*-diphenyl-[1,1'-biphenyl]-4,4'-diamine [65181-78-4]; (b) 4-(diethylamino)benzaldehyde diphenylhydrazone [68189-23-1]; (c) 4-[2-[5-[4-(diethylamino)phenyl]-4,5-dihydro-1-phenyl-1*H*-pyrazol-3-yl]ethenyl]-*N,N*-diethylbenzenamine [57609-72-0] (5-(*p*-diethylaminophenyl)-1-phenyl-3-(*p*-diethylaminostyryl)pyrazoline); (d) 4-(5*H*-dibenz[*a,d*]cyclohept-5-ylidenemethyl)benzenamine [119344-18-2]; (e) 4,4'-cyclohexylidenebis[*N,N*-bis(4-methylphenyl)]benzenamine [58473-78-2] (1,1-bis(di-4-tolyl-aminophenyl)cyclohexane); (f) 4,4'-(phenylmethylene)bis[*N,N*-diethyl-3-methyl]benzenamine [15008-36-3]; (g) 4-(2,2-diphenylethenyl)-*N,N*-bis(4-methylphenyl)benzenamine [89114-91-0]; (h) 4-(9*H*-fluorene-9-ylidenemethyl)-*N,N*-bis(methylphenyl)benzenamine [115838-15-8]; and (i) *N*-[(9-ethyl-9*H*-carbazol-3-yl)methylene]-2,3-dihydro-1*H*-indol-1-amine [87866-87-3].

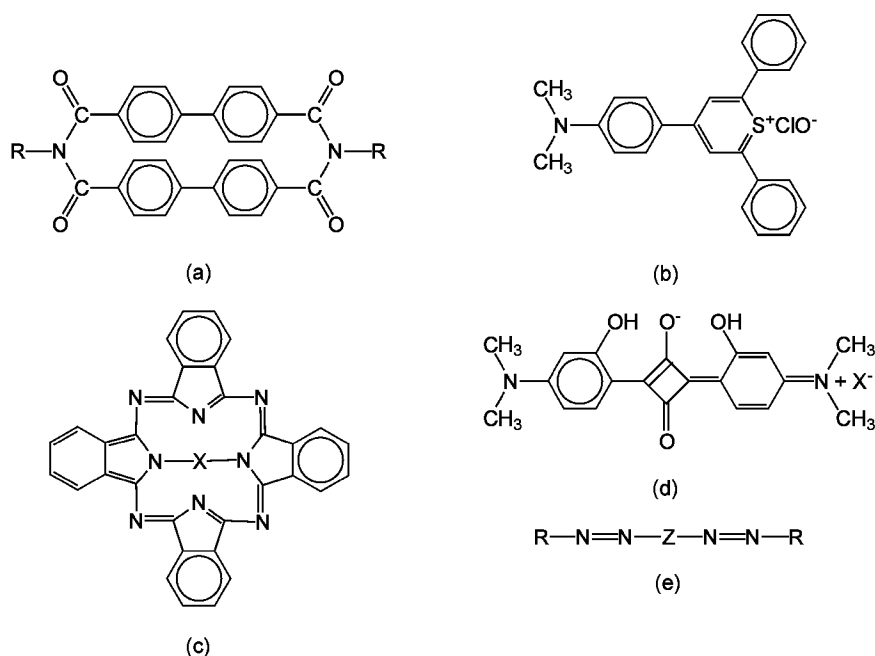


Fig. 6. Classes of charge-generating dyes and pigments: (a) perylenes, (b) thiapyrylium, (c) phthalocyanines, (d) squaraines, and (e) azo. R and Z are organic groups from small alkyl to very complex; Z is typically aromatic.

Conceptually similar are single-layer photoreceptors such as PVK filled with a photoconductive pigment, such as *N,N'*-di(3,5-dimethylphenyl)perylene-3,4,9,10-tetracarboxylic acid diimide (32), or a photoconductor containing a bisazo photoconductive pigment, a hole transporting molecule such as an indolinyl hydrazone derivative of carbazole shown in Fig. 5i, in a polyester binder (33). In practice, this photoconductor is charged positively even though both holes and electrons are mobile.

All the organic photoconductors discussed are single-layer structures the thickness of which is determined by the desired development field, by the dielectric constant of the material, and the resistivity of the device. Organic photoreceptors are typically capable of withstanding higher electric fields than those based on inorganic materials, with the exception of amorphous silicon. Thus single-layer organic photoreceptors are usually thinner (typically 10–20 μm) than the original α -Se based photoreceptors (40–70 μm).

Most of the currently used OPCs are, however, two-layer photoconductors schematically illustrated in Figure 7. In these devices, the charge-photogeneration function is separated from the charge-transport function (34). This separation enables independent optimization of each layer. The charge-generation layer (CGL) is optimized for the spectral response and the carrier-supply efficiency (sensitivity), and the charge-transport layer (CTL) is optimized for wear resistance, high charge carrier mobility, flexibility, corona resistance, etc. The CGL is usually thin, between 0.3 and ~ 3 μm , ie, only thick enough to absorb most of the incident light. The CTL is typically 10–25 μm thick. The charge-generation layers are typically sandwiched between the conductive substrate, generally in drum or belt form, and the polymeric charge-transport layer which protects the CGL from the corrosive and abrasive xerographic environment.

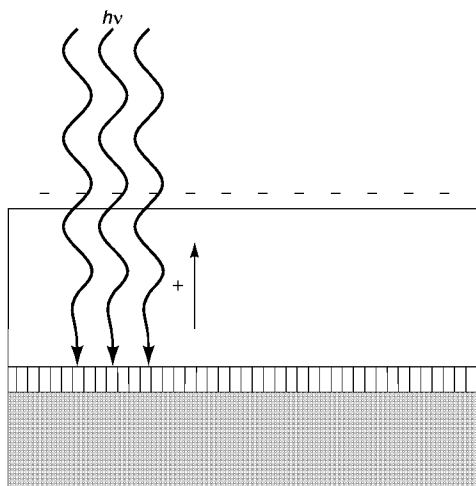


Fig. 7. Schematic of an organic layered photoreceptor, where the – signs represent the corona-deposited charge, which is typically negative; □, the CTL; ▨, the CGL; and ▩, the conductive base. Terms are defined in text.

Most of the known charge-transport layers are *p*-type or hole transporting. Thus this type of layered photoconductor must be charged negatively.

The CGLs contain highly absorbing photoconductive pigments, either dispersed in binder resins or vapor-deposited (sublimed) directly onto the conductive substrate. Photogenerating pigments and dyes that are most frequently mentioned in the literature are the phthalocyanines (35–38), perylenes (39), squaraine and azo-type pigments (40), pyrylium and thiopyrylium dyes (qv) and pigments (41), such as those illustrated in Figure 6, and inorganic pigments such as trigonal selenium in conjunction with an organic CTL (42). A review of the patent literature suggests that, in principle, all dyes and pigments that are known to be photoconductive, are capable of injecting photogenerated charge carriers to typical charge-transporting molecules and polymers. For imaging processes that employ visible, broad-spectrum light sources, the most popular are the bisazo dyes and pigments, primarily because these are relatively easy to synthesize, sufficiently sensitive, and easy to fabricate into photoreceptors. The two-layer photoreceptors may be fabricated by successive solvent coatings of CGL and CTL, although in some cases the CGL pigment is deposited directly by evaporation (see also AZO DYES).

Manufacturing processes differ according to the intended use of the photoreceptor. Drum photoreceptors are typically fabricated by a dip-coating technique where the drum blank is immersed in a solution containing the desired components and slowly pulled up from the solution through a concentric ring. The space between the inner circumference of the ring and the drum surface determines the thickness of the solution layer remaining at the surface and thus the dry layer thickness. The excess solution is brought back into the container. An alternative technique successfully used in photoreceptor

manufacturing is spray deposition, which is basically the same as conventional spray painting used in car body shops, except for the precision involved. The flexible belt photoreceptors are fabricated by gravure or die casting of the solution or dispersion onto a long web of the metalized polyester substrate which, after extensive drying is sliced into segments that are then ultrasonically welded into continuous loops (see also COATING PROCESSES).

Charge-Generation Layers. For xerographic laser printers, photoconductors must respond to the wavelengths of coherent light emitted either by He–Ne lasers at 632.8 nm, or by solid-state Ga–As lasers which emit in the near infrared (ir) region, between 800 and 830 nm. Four groups of compounds that show the near infrared response are derivatives of squaraine (43,44), phthalocyanines (38,45), bisazo and trisazo pigments (44), and thiapyrylium (46) or pyrylium pigments (47). Among these groups of materials, phthalocyanines are most popular because of high sensitivity. Particularly sensitive appears to be the Phase 3 titanyl phthalocyanine (48) which requires only $\sim 0.25 \mu\text{J cm}^{-2}$ of light energy at $5.3 \times 10^5 \text{ V cm}$ to discharge the photoreceptor to one-half of the original potential of 800 V. This is roughly the same as the sensitivity of $\alpha\text{-Se}$ in white light. The issues plaguing ir-sensitive photoreceptors are not the lack of sensitivity, but rather the thermodynamic stability of the most sensitive form, excessive dark conductivity, and the stability of the xerographic performance in general.

The sensitivity of a photoreceptor depends to a large extent on the method of pigment preparation, however, and factors such as the particle or crystallite size, the nature of the binder polymer, the CGL thickness, the pigment concentration, the method and time of milling the pigment particles, the applied electric field, temperature, and even the nature of the adjacent CTL can come into play. No direct comparison of photoreceptors is possible without specifying all the fabrication and material parameters and the testing conditions. The only meaningful way of comparing various kinds of photogenerating pigments would be a parallel test of fully optimized photoreceptors. Such information is not available.

The design of photogenerating materials is still largely empirical but some tentative design criteria appear to be emerging. The CGL must typically be in an aggregated or crystalline form to have high absorptivity and high photogeneration efficiency; have recognizable donor and acceptor groups separated by a conjugated system of double bonds or an aromatic structure to support polarization of the excited state; have reversible redox for stable operation; and have a broad absorption spectrum for light-lens imaging, or, if near ir-sensitivity is desired, must have extended π -conjugation.

Charge-Transport Layers. The charge-transport layer materials are largely responsible for the overall mechanical characteristics of the photoreceptors. These can be classified into four principal classes: (1) polymers having pendent transport-active groups, such as PVK, where the main chain can be vinylic, acrylic, or similar, and the active pendent group is typically a complex aromatic amine or a large polynuclear aromatic group such as anthracene; (2) molecularly doped polymers, which are essentially solid solutions of transport-active molecules (20–50 wt %) in tough polymers such as polycarbonate or polyester; (3) condensation polymers having transport-active groups acting as building blocks of the polymer chain (49); and (4) polymers having a transport-active chain or at least segments of the chain, such as π -conjugated or σ -conjugated polymers represented by polyphenylenevinylene or polysilylenes (50). Examples of these classes of hole-transporting organic media are illustrated in Figure 8.

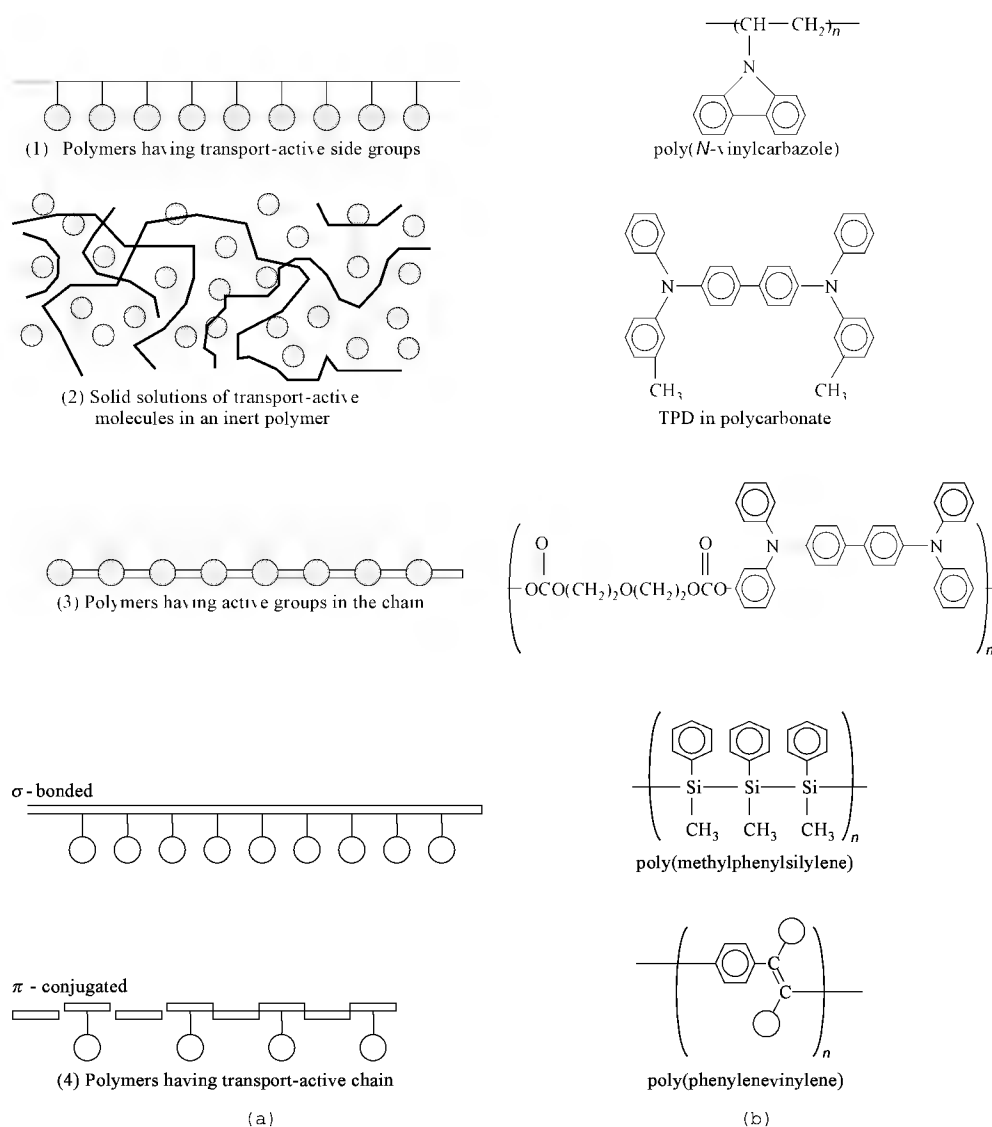


Fig. 8. (a) The four basic classes of charge-transporting polymers, and (b) corresponding examples. Class 4 polymers may be either σ -bonded or π -conjugated. Only classes 1 and 2 have been used in copiers and printers.

Even though numerous patents describing all four classes have been issued, almost all existing OPCs now in use in copiers and xerographic laser printers employ the molecularly doped polymeric charge-transport layers. The transport-active molecules are typically complex aromatic amines having strong electron-donor character (see Fig. 5). It is commonly accepted that these materials transport charges by a hopping mechanism involving charge transfer among discrete sites. These sites are either discrete molecules, pendent groups on the polymer backbone, or segments of a transport-active chain. The hole-transport process is a sequence of redox exchanges among these neutral, uncharged species and the charged, electron-deficient counterparts or cation-radicals. The primary prerequisite is complete reversibility of the redox process. The oxidation of transport-active species, ie, transfer of a hole, to form a cation-radical, and its subsequent reduction back to a neutral molecule must be free of any chemical side reaction which would inevitably result in an immobilization of the transiting hole. Aromatic amines are rather special in that typically the redox processes are completely reversible.

Aromatic amine-based transport materials transport only holes; thus the corresponding photoreceptors must be charged negatively, necessitating positively charged toners. Toners are needed to develop the negative latent electrostatic image which remains on the surface after the exposure. Remarkable progress has been made in triboelectrification of toner particles by using special charge-directing additives and by the appropriate selection of the toner resin. When toner charged to one polarity is used to develop the latent image created by the opposite charge on the photoreceptor surface, the process is called charged area development (CAD). The advancement of xerographic laser printing, however, precipitated discharged area development (DAD), a process in which the toner develops only those areas that lose charge. Both processes are shown schematically in Figure 9. The main advantage of the DAD process is that it is less wasteful: only a fraction of a typical copy area is covered by black marks, and it therefore makes sense to discharge only those areas that have to carry the black marks. Also, the sharpness of the image may be improved because the toner, which is charged to the same polarity as the unexposed areas on the surface, cannot easily spill into those areas still containing charge.

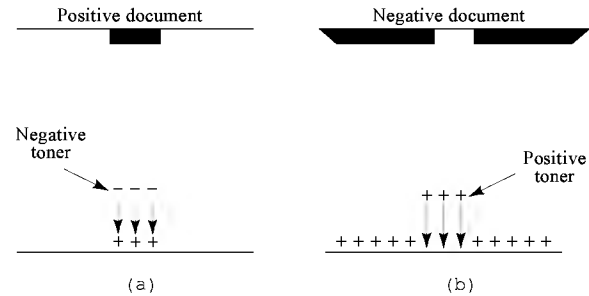


Fig. 9. (a) Schematic of charged area development (CAD), using toner charged oppositely to that of the photoreceptor and resulting in a positive document; (b) discharged area development (DAD), where the toner and photoreceptor polarity are the same, resulting in a negative document.

There are two ways to enable use of DAD with organic photoreceptors: one is to use negatively charged toners and photoreceptors, the other to use positive toners and positively charged photoreceptors. Positively chargeable photoreceptors would either have to use electron-transport layers, or their structure would have to be inverted, with the sensitive photogeneration layer facing the environment and a hole-transporting layer underneath. A variety of electron-transport materials have been described, some having respectable charge mobilities, but as of this writing none is used in commercial xerographic photoreceptors.

The trend in the xerographic copying and printing industry is to expand the use of organic photoconductors which are typically inexpensive and have a useful life. However, the α -Si:H photoreceptors, even though more expensive to manufacture, may ultimately have an advantage in greater durability.

The art of making photoconductors for electrophotography is fairly mature and the utilization of light energy is almost complete at any wavelength. The quantum efficiency of charge generation and collection approach unity, the charge-transport materials move charge without trapping, the action spectrum can be tuned to match the wavelength of the light source, and the materials are mechanically durable. Improvements would involve materials or device designs that retain these properties but produce further reductions in cost of ownership resulting from extension of useful life and alternative, simpler, and ecologically compatible manufacturing processes.

Developer Materials

The Development Process. In the original electrophotographic demonstration, development was accomplished by dusting lycopodium powder over an exposed sulfur film. This yielded low density images of poor resolution. Considerable powder settled in the exposed background areas (the white areas of a document), and image transfer to paper could only be achieved by prior coating of the paper with wax or another sticky material. Therefore, the development of organic polymer materials was of critical importance to the success of xerographic technology (51).

The physical basis of development, which may be simple in principle, is complex in practice (3). A uniform surface charge on a photoreceptor of infinite extent produces associated electric fields, the lines of force of which are confined within the photoreceptor, because these lines terminate on the induced charges of opposite sign in the conductive substrate (Fig. 10). This is not true if the photoreceptor has finite dimensions or if a nonuniform distribution of surface charge exists, because discontinuities exist in the surface charge density. At these discontinuities, force lines extend out of the photoreceptor into space. These so-called fringe-fields, associated with nonuniform distribution of surface charge produced by the imagewise exposure of a photoreceptor, though quite strong, are confined to a narrow region at such a boundary. External to the photoreceptor, these fringe-fields determine the development field E_D and the electrical force, $F_D = Q \times E_D$, on a toner particle having total charge Q . Because such fields at the center of solid areas are very small, fringe-field development is such that only narrow lines or edges of large-area black images are adequately developed (2–4). This problem was overcome by using a biased-development electrode positioned close to the photoreceptor surface so that the lines of force are pulled out of the photoreceptor, even if the receptor is uniformly charged.

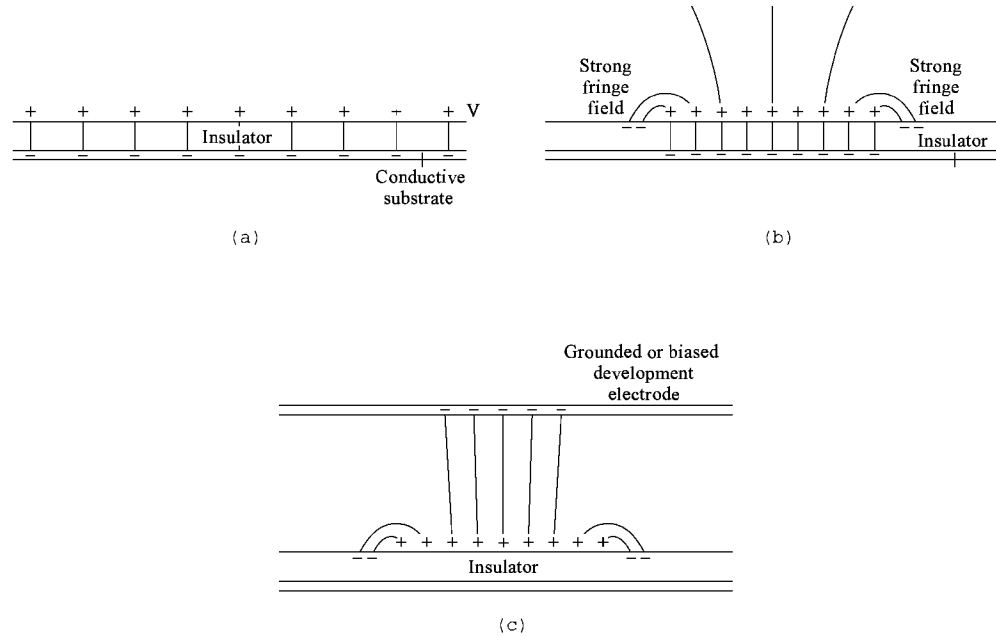


Fig. 10. The lines of force above the photoreceptor for three types of electric fields: (a) zero field; (b) weak field, where a nonuniform charge pattern is produced by selective photodischarge and the field on charged toner particles is strongest at the edges of the charge distribution, constituting the fringe fields; and (c) strong field where an electrode is placed above the charged photoreceptor surface, enhancing the field strength in the center of a broad image.

The development fields, which fall off rapidly over distances of micrometers moving away from the photoreceptor surface, exert an electrostatic force on a triboelectrically charged particle of these dimensions. The average toner particle is only 10 μm in diameter, and the challenge is to control and direct large numbers of such particles. The coulombic forces between charged toner and the charges on the photoreceptor surface are used, but to come under the influence of these electrostatic fields the toner must be brought within micrometers of the photoreceptor surface.

A variety of methods can be employed to achieve the required proximity, most involving an intermediary component called a carrier. These different development processes include powder-cloud, cascade, and magnetic-brush development (2,3). Liquid development, in which an insulating hydrocarbon liquid is used as the carrier, allows the use of much finer toner than can be typically controlled as a dry powder and therefore inherently offers the possibility of much higher resolution development. Powder-cloud or aerosol development uses air as the carrier. Shortcomings arise from poor control over the sign and magnitude of the charge on the toner, leading to excessively dark backgrounds. Additionally, in powder-cloud development, as the name implies, a cloud of toner particles is responsive to air movement, inherently leading to a dirty environment in a xerographic machine. There is also the inability to manage the majority of the toner cloud, because it is only the toner within micrometers of the photoreceptor surface that experiences the development fields. For certain specialized applications, however, such development can be the process of choice. One example is in xeroradiography where small changes in transmitted x-ray intensity need to be detected. Without a development electrode the xerographic images give much higher image contrast in edges, as compared to conventional radiology, thus enhancing discontinuities such as fractures. Using a development electrode, high resolution, continuous-tone xeroradiographic images can be obtained that show particular benefits in mammography and provide a sensitive screening technique for breast cancer (2).

Other processes use solid carriers for toner delivery, so that the developer consists of both toner and carrier, and is known as two-component development (2,3). The process called cascade development is characterized by a dependence on gravitational forces for its operation. In the predominantly used development process known as magnetic-brush development, the carrier is made of a magnetic material such as iron, so that magnetic forces are the dominant delivery force. A single carrier particle or bead, at about 100 micrometers in diameter, is much larger in size than typical toner and delivers the toner to the surface of the photoreceptor. It also controls and determines the sign and magnitude of the toner charge, for which purpose it is usually coated with a thin polymer film. This coating also provides adhesion for the toner particles to the carrier, although the adhesive force must not be so large as to cause the adhesion to be permanent (Fig. 11). The competing electrostatic, adhesive, gravitational, or magnetic forces must therefore be delicately balanced.

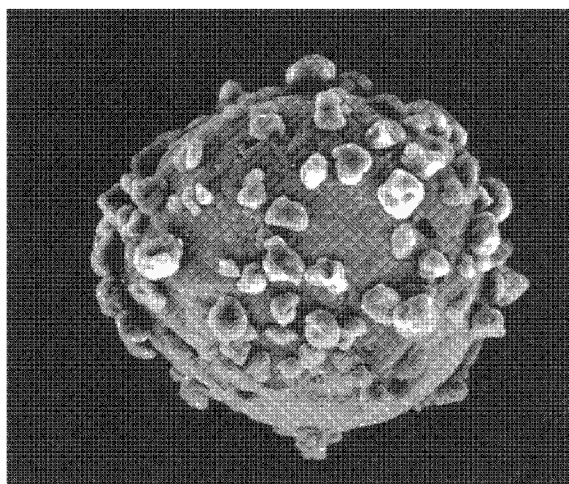


Fig. 11. Magnified photograph of a single carrier bead, 100 micrometers in diameter, showing smaller, attached toner particles.

Courtesy of Xerox Corp.

Triboelectricity. For development to occur, the toner particles must be reproducibly charged to the correct level and polarity for the specific photoreceptor. The phenomena of triboelectricity, which involves the transfer of charge from one solid to another, are exceedingly complex, involving the surfaces of solids and interaction of the surfaces with each other and with the ambient (52). Consequently, the specific experimental observations are highly sensitive to the nature and purity of the materials, the physical and chemical state of both surfaces, and the precise details of the experiments performed. No absolute correlations between theoretical predictions and experimental observations exist.

Triboelectricity originates from the Greek word meaning "to rub," implying that rubbing is essential to the mutual charging of two solids. Rubbing may only serve to provide an area of contact between the solids, however, and the contact may be the true means of charge transfer. Relative roles of rubbing and contact are not easily separated. Whereas some experiments suggest that charging can occur by contact alone, as demonstrated by experiments where one of the materials is liquid mercury (53), it is likely that both contact and frictional rubbing can be important. Empirical correlations between material properties and corresponding triboelectric behavior have allowed a considerable phenomenological base of understanding and operational control of developer materials to exist.

Triboelectricity is a relative phenomenon and, as a result, use is made of ordering the relative charging capabilities of materials. Thus, in the triboelectric series, window glass > polyamide > common salt > wool > silk > cotton > styrene-butadiene copolymer > poly(vinyl butyral) > epoxide resin > natural rubber > sulfur > polyethylene > poly(vinyl chloride) > polytetrafluoroethylene (52-54). Materials at the beginning of this series charge positively with respect to materials near the end. Such a series is used only as a general guide, and the results of any particular combination may deviate from expectations because of contamination, impurities, or humidity variations. A number of efforts have been made to correlate this series with other material parameters. Because the charge exchange may involve the transfer of ions, electrons, or macroscopic amounts of charged material, either separately or in combination, a number of relationships have been claimed. For example, it has been claimed that materials of higher dielectric constant become positively charged when contacted with those of lower dielectric constant (54). Such a correlation is consistent with a charge transfer of weakly bonded ions between solids, because the relative binding forces of these ions to their respective surfaces are reduced through the dielectric constant. Another correlation involves the differential work function, ie, the relative difference in electron energy levels in materials, because charge transfer requires the movement of electrons from one material to another.

Development Process Parameters. If a photoreceptor 50 μm thick is charged uniformly to a surface potential of 1000 V, assuming a typical dielectric constant of 5, the amount of charge per square centimeter of the photoreceptor is 100 nC. One electronic charge amounts to 1.6×10^{-19} C; thus 6×10^{11} singly charged ions would reside on each square centimeter of surface. To achieve complete photodischarge, therefore, 6×10^{11} electronic carriers must be generated for each square centimeter and for a photogeneration efficiency of unity this requires an exposure of 1

$\mu\text{J cm}^2$ (10 erg cm^2).

The photoreceptor is always charged to a particular polarity using the appropriate sign voltage in the corotron, so the toner particles must be charged to the opposite polarity in order to produce a black-on-white copy of a black-on-white original. If the input were a white on black, ie, a negative, then toner with the same polarity would be required to give a black-on-white copy. The latter is an example of reverse or discharged-area development, DAD (55). Photoreceptors in different products may be either positively or negatively charged, so appropriately charged toner of either sign must be available. In two-component development the charging in both sign and magnitude is controlled by the nature of the toner and carrier through a special coating. Further, the degree and uniformity of triboelectric charging are important to achieve the most efficient development and highest quality images. The blackness or density of the image is related to the amount of toner which deposits on the photoreceptor. Thus the charge-to-mass ratio for the toner is an important parameter. The toner-carrier interactions lead to a mutual charging of from $5\text{--}20 \text{ nC cm}^2$, corresponding to about 10^5 electronic units of charge per toner particle or a charge-to-mass ratio of $10 \text{ }\mu\text{C g}$ of toner. For a typical toner, a deposition of about 1 mg cm^2 of toner gives the dark image representative of a high quality copy. If the charge-to-mass ratio is $10 \text{ }\mu\text{C g}$, then the amount of toner needed to give a sufficiently dark image neutralizes 10 nC cm^2 . In the above example 10 nC cm^2 is only about one-tenth of the charge density on the photoreceptor and sufficient development can take place without substantially perturbing the development field (55).

The very small size toner particles required are produced by a grinding or milling process which involves the continual fracturing of polymeric materials. Although uncharged on the average, fresh toner possesses essentially equal numbers of positive and negative charges, as illustrated by Figure 12. The number of particles having a given charge, mass, or charge-to-mass ratio can be measured by a charge spectrometer, and measurements on freshly made toner show that the charge distribution is indeed symmetrically distributed (56). The situation changes when the toner is mixed with carrier beads. The average charge distribution then becomes nonzero, and the sign of the newly equilibrated developer charge depends on the choice of the materials for the component toner and carrier or any additives. The magnitude of the charge also depends both on the time of mixing and the ratio in the mix of toner to carrier (also termed the toner concentration, C_T , and defined as the toner mass divided by the sum of the toner and carrier masses). The average charge grows approximately in an inverse exponential manner with time of mixing and saturates to a value determined by the toner concentration. The inverse exponential growth is consistent with a charging mechanism involving contact in which the fraction of uncharged area on a carrier bead decreases with time. This reflects the fact that the charge density per unit area of a toner particle depends on the probability of finding an uncharged area on the surface of a carrier. This becomes increasingly difficult as C_T increases. Based on this intuitive picture, the charge-to-mass ratio of the toner is expected, and found, to decrease with increasing C_T .

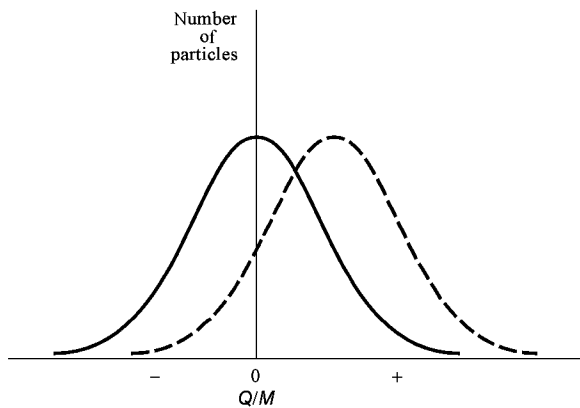


Fig. 12. Charge-to-mass, Q/M , distribution of individual (—) toner and (---) toner-carrier particles produced by milling. Particles are initially positively or negatively charged and, on balance, the net toner charge is zero.

As a result of carrier-coating abrasion, fracturing of carrier beads, and compression of toner particles, considerable complications can occur (3). Carrier beads are commonly coated with low surface energy polymer coatings for control of charge and adhesive forces. Exposure of the underlying carrier beads which may have relatively high surface energies means that any compressed toner can become permanently bonded to them. In such a case, the carrier now presents a toner surface to the other toner particles, so that on the average no charge exchange can take place at those points. This is but one of many examples of aging phenomena that can occur with developer materials and which reveal themselves in progressive loss of image quality, ultimately requiring replacement of the developer.

Charged toner particles attached to carrier beads come into close proximity to the photoreceptor surface. The toner is bound to the carrier by adhesive forces, an electrostatic force F_E and a dispersion force F_A , the combined effect of which must be overcome by the development force exerted on the toner by the latent charge pattern on the photoreceptor (Fig. 13). Because the toner acquired its charge from the carrier, the carrier has a charge of opposite sign and the pair is physically bound by the mutual electrostatic attraction, F_E ; the electrostatic force, F_E , binding a spherically shaped toner-carrier pair is proportional to Q^2/r^2 , where Q is the assumed uniformly distributed charge on both the toner and the carrier and r is the radius of the toner. A single toner particle typically carries a charge of about 10^{-14} C , and F_E is of the order 10^{-8} N , so the electrostatic force on a toner particle is extremely small.

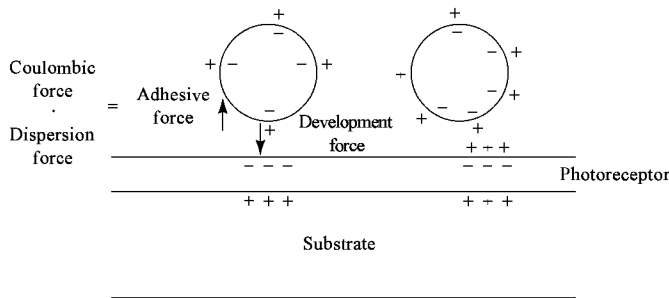


Fig. 13. Two-component development involving competition between electrostatic and adhesive forces, where + outside and beneath the large circles represents the positively charged toner and the large circles containing negative signs represent carrier beads.

The dispersion force, F_A , is a form of van der Waals bonding associated with the interaction of fluctuating dipoles generated by otherwise neutral atoms (3). This force also depends on the separation of the two interacting atoms, but in a way distinctly different from the inverse square relationship of the electrostatic force between two charges. It takes a very short, but finite, time for the electromagnetic signal from one fluctuating dipole to reach a

second. In response to the arrival of this signal, the second atom polarizes and emits its own dipole radiation. If the electromagnetic signal from the second atom is felt back at the first atom before the latter changes its original dipole configuration, then the interactive force is said to be in phase and is attractive in nature. The requirement is that the total time involved, the round trip transit of an electromagnetic signal between the two separated atoms, be less than the period of oscillation of the fluctuating atom. For atomic separations less than about 10 nm, this condition can be met and an order of magnitude estimate of the dispersion force, F_A , between a typical toner-carrier pair gives a value comparable to or higher than the electrostatic adhesive force. For distances substantially >10 nm, any phase correlation is lost and the attractive character of the dispersion force decreases. Thus the strength of the adhesive force is substantially reduced if the toner particles can be separated or detached, albeit slightly, from the carrier beads. Thus, agitation is a very important factor in the development process (3).

Development therefore depends on whether the development force F_D exceeds the forces $F_A + F_E$ which keep the toner attached to the carrier. The electrostatic adhesive force is proportional to Q^2/r^2 , whereas the dispersion force F_A does not depend on the charge on the toner, and for a given set of materials is constant. Hence for a given development field the charge on the toner may be so small that the development force is less than $F_A + F_E$ and development does not occur. However, as the charge on the toner is increased or the toner radius is made smaller, F_E grows faster than the development force and again development does not occur. Thus there is an optimum range of values for F_E , and hence for charge-to-mass ratio, for the toner.

Cascade Development. In cascade development, spherical or near-spherical beads are used to carry finely divided toner powder to the imaging surface (57). Advantage is taken of the gravitational force on the much more massive carrier particle to cascade it, together with the attached toner, over the photoreceptor surface. In the consequent tumbling of the carrier particles across the photoreceptor surface, two effects are significant. In bouncing, the physical impact shakes loose some of the toner which becomes airborne. Because the physical impact has already supplied some of the stripping energy, these particles develop out at relatively low development fields. Toner particles remaining bound to the carrier beads require higher fields for development. Within a few micrometers of the surface, the toner-carrier system feels the effects of the development fields. This force ultimately wins and the toner particle moves toward and becomes attached to the photoreceptor. The retention forces between the toner and carrier in two-component developers give such systems a higher contrast because of a threshold effect. Unless the development fields exceed the forces retaining the toner on the carrier, no deposition occurs. This contributes to clean, dust-free background on nonimage areas.

The earliest copiers used an insulating carrier of sand and had as a principal shortcoming the inability to reproduce solid black areas because development was only by fringe fields. Improvements were achieved by using a closely spaced development electrode or conductive steel shot as the carrier. The conductive carrier itself can serve as a development electrode. However, for a given carrier bead material, the gravitational force (equal to the carrier weight) is proportional to the cube of the bead radius, whereas the carrier charge, related to the charge per unit area, is proportional to the radius squared. Unfortunately, carrier particles have a distribution in sizes so that small carrier particles can be culled out of the cascade flow and become attached to the photoreceptor because the electrostatic forces on them can exceed those resulting from gravity. In addition to possible physical damage to the photoreceptor surface, this limits the achievable flow in cascade development and was a constraint for higher machine speeds. To overcome this problem, carriers were made out of magnetic materials (qv) and, by using magnets, the resulting magnetic fields prevailed over gravity. Most modern xerographic machines use magnetic carriers and magnetic-brush development.

Magnetic-Brush Development. In magnetic-brush development, the carrier is made of beads of ferromagnetic material coated with a thin polymer charge-control film. Ferrite carriers are favored because these are lighter than steel and have a much lower residual magnetic field, allowing ferrites (qv) to mix more readily and flow better in use. The necessary magnetic fields are provided by stationary permanent magnets located within a rotating drum or sleeve which transports the toner-carrier combination to the photoreceptor surface (3). The magnetic fields cause the magnetic cores of the carriers to become individual magnets having north and south poles, as shown in Figure 14. As a result the carrier particles, to which toner is attached, link up to form chains or bristles (hence the name magnetic-brush development), and much higher process speeds can be achieved. The relative speed of the rotating developer drum with respect to the moving photoreceptor plays an important role. At too-high speeds the centrifugal force can result in carrier particles escaping from the constraining magnetic field. If these are stripped of the toner by impacts, then the oppositely charged carrier beads can end up on the nonimage areas of the photoreceptor, an effect known as bead carry-out.

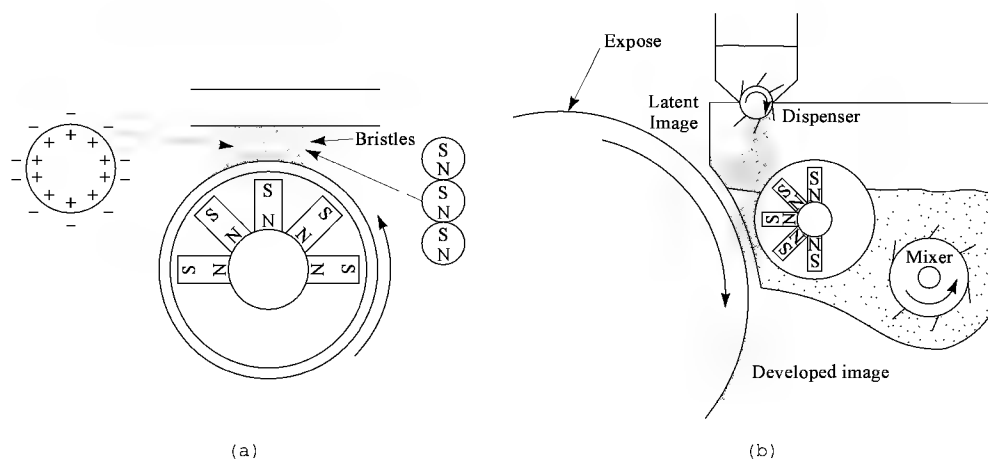


Fig. 14. Two-component magnetic brush development showing (a) the magnetic carrier particles (large circles) carrying toner (-), which within the magnetic field of the rotating permanent magnets, behave as individual bar magnets; and (b) the production of a developed image. See text.

Two types of magnetic-brush development, known as insulating and conductive magnetic brush, are employed. The distinction is based on the effective conductivity of the developer and more specifically the assembly of contacting carrier particles (3). These are controlled by the chemical composition and thickness of the polymer coating of the carrier beads. Conductive magnetic brush gives improved solid-area development because of the reduced effective spacing between the photoreceptor and the developer roller caused by the conductive paths within the developer; in essence, the developer itself functions as a development electrode. As in cascade development, the beneficial impact-associated separation of toner and carrier coexists with the deleterious physical impact and permanent bonding of toner particles to the carrier surface or fracture of the toner or carrier.

One drawback of magnetic toners is apparent when colored developers are needed. Materials having the required magnetism are typically brown to black in color and limit the quality of colors achievable. A more recent change in development systems is the use of single- or monocomponent developers. As the name implies, only one kind of particle is used. This particle serves as the toner and can be magnetic or nonmagnetic depending on whether ferromagnetic material is incorporated. No carrier particles are used and the system is simpler, smaller, and avoids some aging effects. There is another material involved in controlling the necessary triboelectric charging of the toner. This charge control is accomplished by coating a roller with the appropriate material, just as carriers are coated in two-component developers, and the roller thus acts to control not only the charging but also the transport and metering of the toner. This, or a similar roller, can also be appropriately biased to function as a development electrode. A particular case of single-component development is jumping-toner development in which triboelectrically charged toner covering a donor roll rotates synchronously with the moving photoreceptor. By use of electric fields between the donor roll and photoreceptor, selective transfer of toner occurs. The simplicity, conciseness, and low cost of this approach enabled the introduction of throwaway copier cartridges.

Once the latent electrostatic image has been developed on the photoreceptor surface, it must be transferred to plain paper with no loss in clarity.

Figure 15 illustrates this process. Uncoated paper is brought into contact with the photoreceptor without smudging, followed by corona charging using the same polarity of charge (assumed positive) to that used to originally charge the photoreceptor. Prior to the corona charging, the paper seldom makes intimate contact with the photoreceptor except at a few points; elsewhere an airgap separation of several micrometers is likely to exist. On charging with positive ions, negative charge is induced in the substrate and the top surface of the photoreceptor. Under normal conditions of humidity paper is a good insulator and acquires a potential of several hundred volts during this process. As a result, in unexposed areas an attractive field exceeding 10^5 V/cm exists between the paper and the photoreceptor. The pressure from this force pushes the paper into intimate contact with the photoreceptor and ensures toner adhesion to the paper following separation. The transfer efficiency is maximized by the insulating character of the toner and 80–90% of this can be transferred, even at process speeds of two copies per second.

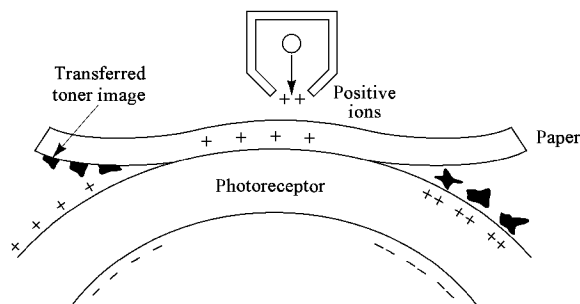


Fig. 15. Transfer of the toned image from the photoreceptor surface to paper is accomplished by charging the paper with a corotron of appropriate polarity.

Toner. Thousands of metric tons of toner, carbon black, and carrier materials are produced and consumed each year having a retail value of several billion dollars. The most widely used toner manufacturing process uses brittle materials that can be milled or pulverized at a high rate, thus increasing production volume and reducing costs. As well as the important triboelectric characteristics, other properties must also be considered. Additionally, the choice of materials must be compatible with volume manufacturing, reproducibility, and cost. Further, the materials must not be toxic, allergenic, or release unacceptable odors or fumes on heating in the fusing step. These materials must be insensitive to humidity and not clump in storage in a machine under a wide range of ambient conditions. The thermal–rheological properties of the toner polymer are critical in determining the fusing characteristics. Moreover, the important flow properties of the toner in the machine are affected by both particle shape and size distribution. Spherical carrier beads have the advantage of flowing more freely so that fresh toner, which must be added from time to time, is charged faster and with greater uniformity. This also prolongs developer life by allowing more uniform surface coatings and minimizing aging effects caused by impacts.

Almost all black toners consist of a thermoplastic polymer or polymer blend with carbon black as the pigment in a weight ratio of about 85% polymer, 15% carbon black. The earliest toner formulations utilized a polystyrene–carbon black mixture. However the properties of the polystyrene led to unacceptable filming of the photoreceptor. This was solved by blending poly(butyl methacrylate), an acrylic resin, with the polystyrene. Styrene–acrylic, styrene–butadiene and polyester polymers and copolymers are largely used for negatively charged toners for use with positively charged photoreceptors such as amorphous selenium and its alloys; polyamides (qv), polyethylene, and ethylene–vinyl acetate copolymers are used as positively charged toners for polymeric-based photoreceptors, which are typically used with negative charge. The carbon black pigment is produced by the burning of oil in specially designed furnaces, which produces the extremely fine particle sizes (on the order of nanometers) required (see CARBON, CARBON BLACK). For color xerography, the three primary colors cyan, magenta, and yellow are needed for process or continuous color (qv) images. A variety of organic pigments exist that provide these colors. However, color is much more subjective than black or white, so that the concentration of the added color calls for particular skill and judgment.

The toner resin and pigment are mixed in either a batchwise or continuous operation using high shear mixers to produce a blend having a taffylike consistency. This is rolled into slabs, cooled, then broken into smaller pieces and fed into a crusher and pulverizer. The resultant powder is then pulverized in a jet mill using high speed air flow in which interparticle collisions and abrasion produces micron-size powder. After screening to remove coarse particles the toner is classified into particular size distribution by a number of techniques. In one of the final processing steps, additives such as metallic stearates, hydrophobic silica, or other materials which enhance flow and cleaning properties are often added. Finally, the toner and developer are comingled so that triboelectric properties such as the charge-per particle and charge-per-mass ratios can be determined. This qualification ensures that the total developer package meets specifications.

Carrier Materials. Carrier composition includes the core material and various carrier coatings that may be employed. A variety of carrier-core materials have been used such as spherical or nonspherical steel shot, sponge or porous iron, and ferrites (58). Processes used to produce carrier materials include atomizing or electrolytic and chemical reduction methods to produce steel or iron carriers, or spraying techniques for ferrites. The carrier beads are typically coated with a thin polymer film to provide a uniform carrier surface, which also serves as a low surface energy layer to reduce impact adhesion of toner to the carrier, thus extending developer and photoreceptor life. These carrier coatings are also used to control the charging characteristics such as polarity and charge-to-mass ratio.

Imaging Media. Various physical properties of paper (qv) are critically important not only in the transfer process, but also in its stripping from the photoreceptor surface, high speed transport through the machine, and in the fusing step. Paper, an interpenetrating network of cellulose fibers, possesses a significant amount of free volume occupied by air, whereas xerographic-grade paper also contains fillers to enhance its mechanical strength and smoothness, and to minimize the shedding of fibers. All of these factors are important in minimizing paper jams. The fillers can control the whiteness of paper, important in determining the image-character contrast density, through its reflectivity. Paper's propensity to absorb moisture because of the hydrophilic or water-attracting nature of cellulose, has considerable implications for the transfer and fusing steps. For example, paper in an ambient relative humidity of 50% can contain up to 8 wt% moisture. This moisture can become the controlling factor in the paper's conductance. As a result, ions from the transfer corotron can leak or flow through the paper thereby decreasing toner transfer efficiencies. Thus paper is an active material in the successful practice of xerography.

Fusing. As the toner image is transferred from the photoreceptor to the paper, the image must be permanently bonded to the paper, rather than attached only by electrostatic and dispersion forces. Otherwise the toner images are easily smudged. In the fusing step, toner particles are made to coalesce with each other, and interpenetrate the paper fibers. The thermoplastic toner may be softened or melted using heat, pressure, or a combination of both. This must happen using minimum energy both for economic reasons and to avoid damaging the paper substrate. The energy needed to soften a thermoplastic is related to rheological properties, specifically the glass-transition temperature, T_g (see RHEOLOGICAL MEASUREMENTS). T_g is the temperature at which the polymer chains of the thermoplastic acquire sufficient thermal molecular motion that it begins to soften, is also characterized by a rapid drop in viscosity through the transition (59). Toners are generally designed to have glass-transition temperatures lying between 50°C and 65°C. Such a narrow window is necessary, because for T_g values less than 50°C unacceptable caking of the developer might occur in storage or the developer housing under ambient conditions. On the other hand, if T_g is too high, excessive amounts of fusing energy are required. The T_g is a function of the polymer molecular weight, and the viscosity of the melted toner is affected by the nature of the polymer or blends, the molecular weight, and the degree of cross-linking within the polymer. Therefore careful consideration has to be given to these thermal and viscoelastic aspects of toner materials. Representative glass-transition temperatures for toners are about 60°C and fusing temperatures are about 130°C. Because the boiling point of water (100°C) is less than the fusing temperature of toner, any water absorbed in the paper is itself vaporized before the fusing temperature is reached. Given the high thermal capacity and latent heat of evaporation of water, a sizable part of the fusing energy can be consumed in the preelimination of moisture (3).

Early fusing systems, such as that used in the Xerox 914, employed radiant heat produced by incandescent lamps. Black toner is an extremely good

absorber of the infrared whereas paper is not (60). Water, having absorption in the infrared, competes with the toner particles for this energy. Radiant fusing has the benefit of instant-on warmup and indirect contact with the toner image. However, the small, but real, potential exists for paper to ignite if it becomes stuck in the fuser system. Another disadvantage is that the photoreceptor, particularly if it is sensitive to red light, must be shielded from stray light from the radiant lamps. Modern copier products use hot-roll fusing in which a combination of pressure, shear force, and temperature are used. The hot-roll is typically a metal drum coated with a silicone rubber. Because the drum operates at a lower temperature, it can handle a wider range of paper stock and can fuse effectively and more efficiently at high speeds. There is no associated visible light. One problem with this system is that of offset, which is melted toner build-up on the hot roller. This is minimized by using a silicone oil release agent on the roller (61) (see RELEASE AGENTS).

An alternative fixing process is cold-pressure fixing. This uses steel rollers of high quality surface finish having nip pressures between the rollers that exceed 18 kN m (100 ppi). The temperature rise associated does not lead to significant softening of the toner so that the fixing is mostly a consequence of cold flow. This means that softer toners than those used in hot-roll fusing must be employed, eg, polyethylene-type polymers. Because the method involves no direct heating, there is no warm-up time and the energy requirements are obviously lower. However, the enormous pressures cause permanent calendering of the paper, in which the thickness of the paper is actually reduced as air is excluded and the fibers compressed. As a result, the toner image typically exhibits an unattractive gloss (60).

The Cleaning Step. Although toner transfer is a highly efficient process, some toner residue remains on the photoreceptor surface and must be removed prior to the next cycle; otherwise successive corona charging and image exposure would be progressively and adversely affected (3). For such toner can capture further toner, and the absorption of visible light could reduce subsequent exposures of the photoreceptor and degrade achievable image quality. The cleaning process employs direct physical contact with the toner and photoreceptor (2). One approach involves the use of brushes; the second uses a blade. Residual toner is bound to the photoreceptor surface by a combination of dispersive and electrostatic forces and the physical forces exerted by the brush or blade must overcome these (62). Because dispersion forces increase with the area of contact, careful design of the preceding steps of development and transfer to avoid the deformation of toner is important in solving this cleaning problem.

In the case of brush cleaning, a number of design factors have to be considered. Among these are the elastic properties and density of the synthetic bristles. Historically rabbit fur worked very well (3). In the case of blade cleaning, different factors must be taken into account. The blade, which contacts the surface of the photoreceptor, must be conformable, whereas the design must take into account that micrometer-sized toner particles must be removed. This obviously requires exceedingly close tolerances over the full length of the photoreceptor. It is for this reason that conformable elastomers (qv) are typically used, although thin, flexible steel blades can be employed for a hard photoreceptor such as amorphous silicon. In either case, the angle of attack of the blade and the frictional forces between it and the photoreceptor must be carefully considered. As before, efficiency of blade cleaning is determined not only by the toner properties but by those of the photoreceptor. The functionality of blade cleaning is highly dependent on the smoothness of the photoreceptor surface because any roughness lifts the blade, allowing toner to avoid removal. The cleaning step involves direct physical contact and puts additional constraints on the photoreceptor.

After the original latent electrostatic image produced by the projection of an optical image on the precharged photoreceptor has been developed by toner, the toner transferred, and residual toner cleaned, there is still some remnant of the original latent electrostatic image. If this remnant is not to interfere with the next image, it must be erased; otherwise the phenomenon of ghosting occurs in which a previous image subsequently prints out. Removal of the latent image is usually achieved by uniformly exposing the photoreceptor to light from an erase lamp which photodischarges the entire surface of the photoreceptor.

Summary

Organic photoreceptors predominate in electrophotography because these materials offer a large degree of design flexibility at low cost. The unit manufacturing costs of α -Si:H photoreceptors are significantly higher than those of organic photoreceptors. Plasma-deposited α -Si:H photoreceptors, which have some shortcomings in these respects, have nonetheless entered the commercial marketplace because of demonstrated durability and long life (63). Photoreceptor life is an ill-defined term, unless described within the context of a specific machine design. As an example, in the case of high speed copiers where, in one engineering approach, a flexible belt photoreceptor employs full document exposure using a light pulse, no matter what the advantages of amorphous silicon-based materials, the high internal stress characteristics of the materials are incompatible with production in the form of a flexible device. Thus α -Si:H photoreceptors are constrained to situations where drum configurations are architecturally acceptable and mechanical integrity and stability are inherently maximized. Such requirements are particularly critical in high speed, high volume printers, including digital laser copiers.

In the field of toner materials, the emphasis in the 1990s is on increasing developer life, particularly using color toners. The life of a typical black two-component developer is about 80–100 kilocycles per kilogram, whereas for color toners it is only about 10 kilocycles per kilogram (64). Also, there is a need for better control of triboelectrification of toner particles that would not only allow stable operation under all environmental conditions but also provide consistent, fast, and reproducible contact charging.

The trend in the reprographic business, not only in printing but also in copying and duplicating, is towards the digital processing of information, rather than using analogue light-lens reproductions. Therefore, photoreceptors have to respond to sources of digital input. This can be either radiation emitted by lasers, light emitting diode arrays, or radiation filtered by fast-responding liquid crystals. Because the least expensive and best field-tested lasers are solid-state GaAs lasers that emit in the near ir (680–830 nm) range, photoreceptors must respond to the incident wavelengths. Organic photoreceptors are best suited for this function, and the vast majority of small and medium size xerographic laser printers on the market use OPCs. The possible advancement of either inexpensive solid-state lasers emitting in the visible wavelength range or other sources of digital input such as light-emitting diodes, could diminish the need for such near-ir sensitive photoreceptors (see LIGHT GENERATION).

The main factors that limit the speed of duplicators are paper handling and the input of energy needed to accomplish fusing of toner to paper. In the case of laser printers the main factor is the speed of the rotating polygon, the purpose of which is to deflect the laser beam signals to the appropriate spots on the photoreceptor. The core materials, the photoreceptors and the toners, do not limit the technology. For example, a laser printer operating at speeds greater than three standard 20.35-cm (8-in) wide pages per second having the desired print resolution of 1500 lines per centimeter, with a 16 mirror polygon (a hexadecagon) would have to rotate at 54,000 rpm.

Fusing toner to paper requires a significant amount of energy. The faster the throughput of paper, the more energy needed to heat paper and toner to fusing temperatures that are typically in the neighborhood of 100°C. Enough energy could be supplied and stable fusing temperatures maintained, even at very high throughput speeds, provided the flow of paper through the fuser is steady. However, in real situations, copying and printing jobs come in spurts and the fuser rolls may easily overheat when the flow of paper which cools the rolls is suddenly stopped. The fuser-roll materials, the heat-transfer parameters, and heat capacities of all the materials involved, rather than the photoreceptor performance, thus constitute an important limitation on the printing speeds.

Future development trends in xerographic materials is also dependent on the emergence of color xerography. The triboelectrification of pigmented (color) toner particles has to be improved. In many instances, the colorant itself may direct the triboelectrification in the wrong direction and its effect must be overridden by additives or formulations which do not affect the color of the toner. Then, the average toner particle size of about 9–13 μ m, typical of black-only toners, may well have to be reduced to 7–8 μ m. Smaller toner particles enable better copy quality, through reduced image pile height and decreased image graininess (64).

BIBLIOGRAPHY

"Reprography" in *ECT* 2nd ed., Vol. 17, pp. 328–378, by J. J. Bulloff, State University of New York at Albany; "Electrophotography" in *ECT* 3rd ed., Vol. 8, pp. 794–826, by N. E. Wolff, Consultant, and J. W. Weigl, Xerox Corp.

1. J. Mort, *The Anatomy of Xerography*, McFarland & Co., Publishers, Jefferson, N. C., 1989.

2. R. M. Schaffert, *Electrophotography*, Focal Press, New York, 1975.

3. E. M. Williams, *The Physics and Technology of Xerographic Processes*, Wiley-Interscience, New York, 1984.

4. C. F. Carlson, in J. H. Dessauer and H. E. Clark, eds., *History of Electrostatic Recording, in Xerography and Related Processes*, Focal Press, New York, 1965, Chapt. 1.

5. E. C. Giamio, *RCA Rev.* **23**, 96 (1962).

6. **U.S. Pat. 2,221,776** (1940), C. Carlson.

7. **U.S. Pat. 2,297,691** (1942), C. Carlson.

8. J. H. Dessauer and H. Clark, *Xerography and Related Processes*, Focal Press, London, 1965.

9. **U.S. Pat. 2,970,906** (1961), W. E. Bixby (to Xerox Corp.).

10. S. G. Grenishin and I. A. Cherkasov, *Zh. Nauch. Prikl. Fotogr. Kinematogr.* **5**, 433 (1960).

11. J. H. Neyhart, *Photogr. Sci. Eng.* **10**, 126 (1966).

12. K. Kiyota, A. Teshima, and M. Tanaka, *Photogr. Sci. Eng.* **24**, 289 (1980).

13. E. J. Felty, in O. Helwich, ed., *Proceedings of the 2nd International Congress on Reprography, Kyn, Germany, 1967*, Verlag O. Helwich, Darmstadt, 1969, p. 40.

14. R. A. Fotland, *J. Appl. Phys.* **31**, 1558 (1960).

15. D. M. Pai and M. Scharfe, *J. Non-Cryst. Solids* **8**, 752 (1972).

16. J. Young and H. G. Grieg, *R.C.A. Rev.* **15**, 471 (1954).

17. Ref. 2, p. 318.

18. W. Eckenbach, in W. F. Berg and K. Kauffe, eds., *Current Problems in Electrophotography*, deGruyter, New York, 1972, pp. 133–138.

19. R. B. Comizzzoli and D. A. Ross, *Photogr. Sci. Eng.* **13**, 265 (1969).

20. P. M. Cassiers, K. Verhille, and J. van Engeland, *Appl. Opt. Suppl.* 3 (Electrophotography), 83 (1969).

21. Ref. 2, p. 336.

22. K. Nakamura, *IEEE Trans. Electr. Dev.* **ED-19**, 405 (1972).

23. M. J. Mitsui, *IEEE Trans. Electr. Dev.* **ED-19**, 396 (1972).

24. P. Mark, *Photogr. Sci. Eng.* **18**, 254 (1974).

25. K. Yoshida, K. Kinoshita, and T. Kawamura, *Appl. Opt. Suppl.* 3 (Electrophotography), 170 (1969).

26. S. Faria and R. E. Karam, *J. Appl. Photogr. Eng.* **7**, 102 (1981).

27. W. Spear and P. G. LeComber, *Solid State Commun.* **17**, 1193 (1975).

28. I. Shimizu, T. Komatsu, K. Santo, and E. Inoue, *J. Non-Cryst. Solids* **35/36**, 773 (1980).

29. W. D. Gill, *J. Appl. Phys.* **43**, 5033 (1972).

30. P. J. Melz, *Photogr. Sci. Eng.* **21**, 73 (1977).

31. W. J. Dulmage and co-workers, *J. Appl. Phys.* **49**, 5543 (1978).

32. Nguyen Chang Khe, S. Yokota, and K. Takahashi, *Photogr. Sci. Eng.* **28**, 191 (1984).

33. Nguyen Chang Khe and co-workers, *Photogr. Sci. Eng.* **28**, 195 (1984).

34. P. J. Regensburger, *Photochem. Photobiol.* **8**, 429 (1968).

35. **U.S. Pat. 3,357,989** (1967), J. F. Burne and P. F. Kurtz (to Xerox Corp.).

36. **U.S. Pat. 3,816,118** (1974), J. F. Burne (to Xerox Corp.).

37. **Brit. Pat. 1,337,228** (1973), M. Smith, R. Radler, and C. Hackett (to Xerox Corp.).

38. R. O. Loutfy, A. M. Hor, and A. Rucklidge, *J. Imag. Sci.* **31**, 31 (1987).

39. E. G. Schlosser, *J. Appl. Photogr. Eng.* **4**, 118 (1978).

40. P. J. Melz and co-workers, *Photogr. Sci. Eng.* **21**, 73 (1977).

41. **U.S. Pat. 3,615,414** (1971), W. A. Light (to Eastman Kodak Co.).

42. K. Takeda and M. Nagata, *Photogr. Sci. Eng.* **25**, 268 (1981).

43. A. C. Tam and R. D. Balanson, *IBM Res. Dev.* **26**, 186 (1982).

44. R. E. Wingard, *IEEE Ind. Appl.* **125**, 1 (1982).

45. R. O. Loutfy, A. M. Hor, C. K. Hsiao, and A. R. Melnyk, *Fourth International Conference on Non-Impact Printing*, IS&T, Springfield, Va., 1988, p. 52.

46. K. C. Nguyen and D. S. Weiss, *J. Imaging Tech.* **15**, 158 (1989).

47. J. H. Perlstein in D. A. Seanor, ed *Electrical Properties of Polymers* Academic Press, New York, 1982, Chapt. 2, p. 59.

48. S. Takano and co-workers, *Imag. Sci. Technol.* **17**, 46 (1991).

49. J. S. Facci and co-workers, *J. Phys. Chem.* **95**, 7908 (1991).

50. M. A. Abkowitz and M. Stolka, in M. Lewin, ed., *Polymers for Advanced Technologies*, VCH, New York, 1988, p. 225.

51. P. J. Flory, *Makromol. Chem.* **1**, 5 (1986).

52. W. R. Harper, *Contact and Frictional Electrification*, Oxford University Press, Oxford, UK, 1967.

53. J. Lowell and A. C. Rose-Innes, *Advan. Phys.* **29**, 947 (1980).

54. L. B. Loeb, *Static Electricity*, Springer-Verlag, Berlin, 1958.

55. R. W. Gundlach, in J. H. Dessauer and H. E. Clark, eds., *Xerography and Related Processes*, Focal Press, New York, 1965, Chapt. IX.

56. R. B. Lewis, E. W. Connors, and R. F. Koehler, *Proceedings of the 4th International Conference on Electrophotography, Washington D.C., 1981*, Society of Photographic Scientists and Engineers, Springfield, Va., 1982; **U.S. Pat. 4,375,673** (to Xerox Corp.).

57. A. S. Diamond, *Toner: Black and Powder for an Exploding Market*, Diamond Research Corp., Ventura, Calif., 1979.

58. A. S. Diamond and L. O. Jones, *Copier Powders, Metals Handbook*, Vol. 7, 9th Ed., American Society for Metals, Menlo Park, Calif., 1984, pp. 580–588.

59. R. B. Prime, *Photogr. Sci. Eng.* **27**, 19 (1983).

60. P. G. Andrus and F. W. Hudson, in J. H. Dessauer and H. E. Clark, eds., *Xerography and Related Processes*, Focal Press, New York 1965, Chapt. XIV.

61. D. A. Seanor, *Photogr. Sci. Eng.* **22**, 240 (1978).

62. L. Nebenzahl and co-workers, *Photogr. Sci. Eng.* **24**, 293 (1980).

63. J. Mort, in Arthur S. Diamond, ed., *Handbook of Imaging Materials*, Marcel Dekker, Inc., New York, 1991, Chapt. 10.

64. D. A. Hays, *J. Imaging Tech.* **17**, 252 (1991).

Milan Stolka
Joseph Mort
Xerox Corporation

ELECTROPLATING

Electroplating is the process of applying a metallic coating to an article by passing an electric current through an electrolyte in contact with the article. The ASTM (1) adds some quality restriction by defining electroplating as electrodeposition of an adherent metallic coating on an electrode such that a surface having properties or dimensions different from those of the basis metal is formed. Two other related plating processes are electroforming, an adaptation of electroplating wherein the deposit is removed from the basis material for use as a separate part, and autocatalytic plating, more commonly referred to as electroless plating (qv). This latter is a form of metal plating wherein dissolved metal is reduced to its metallic form on an article using a chemical reducing agent instead of externally applied electric currents, yielding plate thicknesses of any desired level. In immersion plating the plate is very thin, and the replacement reaction ceases when the active surface is completely coated (see also METALLIC COATINGS; THIN FILMS).

The science of electroplating technology has followed the art. As late as the 1930s, platers depended on unpublished trade secrets for much of their craft. Solution formulas often were contained in the platers' jealously guarded little black books. Plating solutions of that period often produced dull, rough, pitted deposits that needed much hand finishing and buffing. During this time, some materials were discovered that produced improved deposits when added to the plating bath. Many of these secret additives were food products and some of the modern proprietary electroplating additives have been derived from analyses of the active ingredients of these grocery additives.

The first basic book on the principles of electroplating and electroforming (2) was published in the United States in 1924. An electroplaters society, American Electroplaters Society, now known as American Electroplaters and Surface Finishers Society, was formed soon after. Chemical suppliers to the plating industry formed and began developing specialty chemicals and additives designed to improve the quality of the electroplating processes. The specialty chemical and anode suppliers contributed significantly to the advancement of electroplating technology (see also METAL ANODES).

Progress in electroplating is also linked to improvements in materials of construction (see BUILDING MATERIALS, SURVEY), power supplies and other plating equipment, purer industrial chemicals and anodes, and improved analytical test and control methods. The quality of electroplating is dependent on the basis metal surface. Cleaner, less porous castings and better casting alloys, and improved steel (qv) and steel finishes have helped significantly (see also METAL SURFACE TREATMENTS; SURFACE AND INTERFACE ANALYSES).

Materials

Surfaces. Essentially any electrically conductive surface can be electroplated, although special techniques may be required to make the surface electrically conductive. Many techniques are used to metallize nonconductive surfaces. These are well-covered in the literature (3) and can range from coating with metallic-loaded paints or reduced-silver spray, to autocatalytic processes on tin–palladium activated surfaces or vapor-deposited metals. Preparation steps must be optimized and closely controlled for each substrate being electroplated.

Although metals and alloy substrates account for much of the volume in electroplating, there is a large and growing amount of plastic surfaces being plated, both for decorative trim and for electronic shielding applications. On a smaller scale, other materials that are plated include wood (qv), plaster, fibers (qv) and cloth materials, and plant and animal tissue, such as leaves, leather (qv), paper (qv), and seashells.

Electroplated Metals and Alloys. The metals electroplated on a commercial scale from specially formulated aqueous solutions include cadmium, chromium, cobalt, copper, gold, indium, iron, lead, nickel, platinum-group metals, silver, tin, and zinc. Although it is possible to electroplate some metals, such as aluminum, from nonaqueous solutions as well as some from molten salt baths, these processes appear to have achieved little commercial significance.

In addition to the metals listed above, many alloys are commercially electroplated: brass, bronze, many gold alloys, lead–tin, nickel–iron, nickel–cobalt, nickel–phosphorus, tin–nickel, tin–zinc, zinc–nickel, zinc–cobalt, and zinc–iron. Electroplated alloys in lesser use include lead–indium, nickel–manganese, nickel–tungsten, palladium alloys, silver alloys, and zinc–manganese. Whereas tertiary and many other alloys can feasibly be electroplated, these have not found commercial applications.

Composites. Another type of electrodeposit in commercial use is the composite form, in which insoluble materials are codeposited along with the electro-deposited metal or alloy to produce particular desirable properties. Polytetrafluoroethylene (PTFE) particles are codeposited with nickel to improve lubricity (see LUBRICATION AND LUBRICANTS). Silicon carbide and other hard particles including diamond are co-deposited with nickel to improve wear properties or to make cutting and grinding tools (see CARBIDES; TOOL MATERIALS).

Economic Aspects

Electroplating is done both in job shops, where a customer's work is plated, and in captive (in-house) shops. There were reported to be about 7500 plating plants in the United States (4) in 1992. This is a decrease from the ca 12,000 reported by the same source in 1980. The reduction, particularly in the number of smaller job shops, is related to the problems in meeting the waste regulations imposed on plating shop effluents.

The plating industry is relatively small. In the United States, the American Electroplaters and Surface Finishing Society (AESF) had about 9000 members in 1992. No colleges or universities offer full courses specifically in electroplating, although a very few offer a degree in electrochemical engineering. Special courses are available from other sources, however (5), and two technical societies, AESF and American Society for Metals (ASM), offer intensive courses on electroplating.

Many captive-plating shops exist, and the value added by electroplating in captive shops can far exceed plating costs. For example a \$10,000 part, scrapped because of wear or mismachining, may be salvaged using \$10 or less of plated metal. The 10 largest electroplating job shop plants in the United States estimated annual sales totalling \$178–217 million with ca 3100 employees (6). Twenty-nine of the next larger shops reported 3300 employees.

One indicator of business in the plating industry is anode sales to electroplaters; however, anode costs are only a varying portion of the prices used by electroplating shops, because no standard pricing practice seems to exist. The total value of the plating industry is estimated to be in the area of several billion dollars. Job shop plating sales could be about a billion dollars.

Cadmium. In 1989, U.S. consumption of cadmium for coatings was 1474 t (7), compared to 1552 t in 1970, 2089 t in 1979, and 1230 t in 1985. Cadmium plating amounts to about 15° of total cadmium production (see CADMIUM AND CADMIUM ALLOYS). Of the cadmium being plated in 1989, 30° was for automotive parts, over 22° for electronics, and 18° for industrial fasteners. Because of cadmium's high and well-publicized toxicity and very tight waste restrictions, there are considerable efforts to develop alternative materials, and the quantities of cadmium used in electroplating are expected to decrease. The price of cadmium anodes in early 1993 was about \$1 kg.

Chromium. Worldwide consumption for functional uses of chromium is estimated at 13,600 metric tons. From 3630–4080 t of this is used in the United States; Europe is estimated to use about 3600 t; and the remainder is divided among Far Eastern and Third World countries. For functional applications, chromium is used for its hardness and wear properties.

Decorative plating primarily over bright nickel is estimated to consume about 2270 t worldwide. Some 60° of this is used in North America, and about 900 t in Europe. A relatively new method in decorative chromium plating is based on trivalent instead of hexavalent chromium, and this is estimated to total about 10° of the decorative market and is expected to increase. Chromic acid, which cost about \$2.76 kg in early 1993, is the hexavalent

chromium source for plating solutions (see CHROMIUM AND CHROMIUM ALLOYS; CHROMIUM COMPOUNDS).

Copper. Copper (qv) consumed in electroplating and electroforming in the United States and Canada is estimated to be 136,050 t yr including, since 1983, the U.S. Government copper plating of the zinc penny used for coinage. This usage averages about 870 t yr and could range up to 1360 t yr. In early 1993, copper prices were on the order of \$3.65 kg (see also COPPER ALLOYS).

Gold. Used both in decorative and electronic electroplating applications, about 27.5 t of gold were consumed in the United States in 1990 (8) (see GOLD AND GOLD ALLOYS).

Silver. About 72.3 t of silver were electroplated in the United States (8) in 1989. Silver is used in dinnerware and hardware or functional applications. This usage is reported to be fairly stable (see SILVER AND SILVER ALLOYS).

Nickel. Worldwide, nickel used in electroplating has averaged about 63,500 t annually from 1980–1990 (9). The United States uses about 18,000 t yr, and Europe about the same quantity; Japan consumes about 9,000 t, and another 9,000 t is used by the other Pacific rim countries. Canada and South America are reported to use about 4500 t annually. Electroforming applications consume another 4500 t of nickel worldwide. About half of this electroforming is done in the United States and Canada. Nickel deposited from autocatalytic solutions was estimated to account for 1600 t of nickel on a worldwide basis (10) in 1990. Nickel averaged ~\$3.65 kg in early 1993 (see NICKEL AND NICKEL ALLOYS).

Tin. Application of tin on strip steel for can stock has decreased. Nevertheless, tin plating is still done in large volume, and tin plate for can stock was estimated at 11,750 t in 1990 (11). Additionally, 603 t of tin anodes were used for electroplating in electronic applications in the United States in 1990. The use of tin in solder manufacture is reported to exceed that used in can stock (12). The cost of tin in early 1993 was \$7.50 kg (see TIN AND TIN ALLOYS).

Zinc. Suppliers of zinc electroplating chemicals have reported that zinc electroplating volumes continue to grow. More recently, electroplated zinc alloys have been promoted. Zinc was priced at \$1.50 kg in early 1993 (see ZINC AND ZINC ALLOYS).

Other Materials. The amount of plastic substrates plated in 1991 for both decorative and electromagnetic shielding applications was estimated to be about 4.3–5.4 × 10⁶ m² (13).

Uses

Electroplated materials are generally employed for a specific property or function. There is, of course, some overlap; for example, decorative use certainly requires some degree of corrosion resistance. The various usages and the principal plating metals employed are as listed. There are also smaller amounts of other metals and alloys used for specific applications.

Property	Function	Principal plating metals
decorative		chromium, copper, nickel, brass, bronze, gold, silver, platinum-group, zinc
corrosion resistance		nickel, chromium, electroless nickel, zinc, cadmium, copper and copper alloys
wear, lubricity, hardness		chromium, electroless nickel, bronze, nickel, cadmium, metal composites
bearings		copper and bronze, silver and silver alloys, lead–tin
joining, soldering, brazing, electrical contact resistance, conductivity		nickel, electroless nickel, electroless copper, copper, cadmium, gold, silver, lead–tin, tin, cobalt
barrier coatings, antidiffusion, heat-treat, stop-off		nickel, cobalt, iron, copper, bronze, tin–nickel
electromagnetic shielding		copper, electroless copper, nickel or electroless nickel, zinc
paint lacquer base, rubber bonding		zinc, tin, chromium, brass
manufacturing; electroforming		copper, nickel
manufacturing; electronic circuitry		electroless copper, copper, electroless nickel, nickel
dimensional buildup, salvage of worn parts		chromium, nickel, electroless nickel, iron

Decoration. Decorative chromium plating may be the most familiar electroplating application. Colloquially referred to as chrome, chromium finishes are composed of multiple coatings, either nickel or copper and nickel undercoats, having relatively thin (0.2–0.6 μm) coatings of chromium as the top coat. Work by the automobile manufacturing companies, the specialty chemical plating suppliers, and other have led to improvements in the corrosion resistance and appearance of chromium-plated finishes, and in improved processes. Corrosion protection has been extended such that plated bumpers have been found rust-free in some exposure tests after 15 years (14). The improved processes require more complicated plating systems, but these have been adopted where better corrosion protection is a requirement.

To plate a smooth, highly decorative finish on base materials requires special processing. For the roughest surfaces, mechanical finishing such as polishing and buffing are needed. Copper is more easily polished than steel, and because copper flows well with buffing, rougher parts are often copper-plated to substantial thicknesses and then polished and buffed. For smoother surfaces, it is sometimes possible to avoid mechanical finishing. A sufficiently smooth surface may be achieved by the use of adequate thicknesses of copper and nickel deposits from higher leveling plating solutions. The term leveling is applied to that property of a plating process that produces a plated surface smoother than the original surface. Leveling may be positive, the surface gets smoother; or negative, in which the surface gets rougher. The ability to level is a function of the individual metal being plated and the additives, usually proprietary, used in the particular plating solution. In some plating solutions, such as alkaline copper cyanide, leveling can also be enhanced by d-c power manipulations (15). Improvements in the leveling capabilities of plating processes have reduced the need for mechanical finishing, especially when combined with improved casting surfaces. For some applications, the need for polishing and buffing is, however, expected to remain. The quality and performance of the plated surface depends substantially on the finish and quality of the substrate material. Thus the basis material has to be considered an important variable in the plating process.

Other decoratively plated coatings produced in quantity are copper and copper alloys. Copper-plated hardware for cabinetry, lighting fixtures, household trim, and other uses is protected using a top coat of clear lacquer. Copper plating finishes are available with a variety of post-treatments to simulate antique or rustic appearances.

Bright brass plating for decorative purposes is usually produced much like chromium; ie, it is plated over a bright nickel plate. The brass is very thin, amounting to only a flash of as little as 0.05 μm. Thicker deposits tend to be duller. For bright brass electrodeposits of greater thicknesses, proprietary brighteners are added to the brass plating solution, but often thicker deposits require buffing to obtain full color and brightness.

Bronze, copper–tin alloy, electrodeposits can be produced in thicker deposits using proprietary brightening additives in the plating solution, especially over a smooth bright substrate. The brightest deposit with better corrosion resistance is attained when bronze is plated over a bright nickel plate. However, direct plating over steel is not an uncommon practice.

Decorative plating of jewelry using gold and silver is one of the oldest uses of electroplating and the first electroplating process to be patented (16). Bright nickel plate is used as an undercoat, especially for gold-plated jewelry, and the gold can be as thin as 0.05 to 0.10 μm.

Decorative silver plating is used with or without a bright nickel plate as an undercoat. When used on jewelry, plating can be thin, 1 or 2 μm, and requires a clear lacquer top coat to retard silver tarnishing. Decorative silver plating as used in plating flatware (eating utensils) and hollow ware such as bowls, pots, urns, and trays, accounts for considerable quantities of silver. The silver plate thickness on flatware varies according to the specific grades used in this industry, and ranges from about 3.5 to 32 μm (17).

A small amount of the platinum-group metals is also used in jewelry. The high cost of these metals is an obvious deterrent.

Although zinc is not usually considered a decorative plating metal, advances in plating processes have resulted in the production of very bright zinc deposits, matching the appearance of bright nickel–chromium. Bright zinc does not, however, withstand the more severe exposure to corrosive environments. Chromated zinc deposits can also be dyed a variety of colors. These may be decorative, and are given a clear lacquer top coat. Brass-like

colors are being produced on parts used in protected environments, such as inside homes and offices.

Functional Plating. Plating for engineering purposes encompasses all the other purposes for electroplating. Plating for functional properties has grown to surpass decorative applications. Platings for functional properties are extremely varied and are present in almost every industry. Work is plated in a variety of shapes and forms, but the simpler shapes are less troublesome in achieving adequate plate thickness distribution.

Plating Method.

Strips. Materials such as strip steel are plated on machines where coils of steel are unrolled on a continuous, high production basis, fed through a sequence of preparation steps, and into the plating tank(s) on a series of rolls and rollers. Short plating times, high current densities, and relatively thin deposits are the rule. Current is supplied to the moving strip through contact with the rolls entering and exiting the plating tank. The large installations are electrogalvanizing (zinc plating) lines, tin plating lines, and the so-called tin-free lines (a form of chromium). Some strip lines have been started up in several countries to plate zinc alloys, primarily zinc–nickel and zinc–cobalt, which are being promoted for making the body steel for automobiles more corrosion resistant.

Wire. In electroplating wire is uncoiled from spools or reels, passed through the processing and plating steps as individual strands, and then recoiled. Several spools can be plated simultaneously through multistrand machines. More flexible wires can be handled by a multipass arrangement where the plated wire is directed back and forth through the plating tank. Wire has also been plated in loose coils without uncoiling, but the plating quality is less. Wire is plated commercially with several metals. Among these are copper and copper alloys, zinc, iron and iron alloys, nickel and nickel alloys, gold, and silver. Special machines are available to plate wire, and come in many sizes from desk top to shop size models.

Stampings. Stampings, moldings, and castings are usually mounted onto specially designed plating racks to allow rigid and proper positioning of the part. Some plating shops prefer to hang parts on universal hooks, or wire parts onto hooks using copper or steel wire, but this does not allow maximization of plate distribution. The positioning and the spacing of the parts on the rack is an important factor in plate thickness distribution. The plating rack is designed to have sufficient electrical contacts and current-carrying ability to provide current without generating excessive heat from high electrical resistances. The electrical current flow between the part and the anode is such that areas of the part closest to the anode get the most plate, and deeply recessed areas may get little or no plating. Additionally, current attracted to the edges and sharp corners of parts and plate thicknesses is greater in these areas. Plating racks are coated with inert plastic materials except for the necessary electrical contact areas. This prevents waste and allows improved current distribution. Racks can also be provided with inert plastic shields to help direct current flow to desired areas. Racks can be designed to provide auxiliary or intermediate anodes, attached to, but insulated from, the rack. These anodes are used to direct current flow to deeply recessed areas of a shaped part. A rack design and supply industry is active and caters to the plating industry. Properly designed racks are important to production of good quality plated ware.

Barrel Plating. When the parts to be plated are small enough, bulk plating methods may be used. Although some work may be electroplated in dipping baskets, plating barrels are effective. Plating barrels come in many shapes, sizes, and styles. All are capable of turning while going through the process, are shaped to tumble small parts so these are continually mixed, and are equipped with holes to allow for passage of current and exchange of solutions. Plating barrels are made of inert plastic materials with means for electrical contact to the parts and the d-c power source. Sizes range from beaker-size to barrels that hold hundreds of kilograms of parts. Parts most acceptable to barrel plating are those that tumble well and do not stick together or nest. Voltages required for barrel plating processes are higher than those required for rack plating because of the added resistance from the small open-hole area through the barrel.

Brush Plating. When parts are large and only smaller areas of the part require plating brush plating is employed (18). Other terms used for this method include selective, contact, swab, and out-of-tank plating. Specially designed plating tools, which are essentially shaped-anode materials covered with some absorbent material saturated with a plating solution, are used. The anode tool is electrically connected to a d-c power source, and the part to be plated connected as the cathode. The circuit is completed by contacting the wet anode-tool to the part to be plated. Tools can be handheld, kept saturated and in motion, or can be fixtured to do rotating parts such as shafts. This technique finds use in plating of large parts, for touch-up and repair, and has been particularly helpful in the aircraft industry.

Automation. Handling of racks or barrels through the process tanks in a plating line can be done manually, or it can be fully or partially automated. A variety of automated styles are used: a return-type conveyor system transports the racks or barrels from a loading station through the sequence of tanks without backtracking; the conveyor returns through the second half of the process cycle off to the side of the first half so that in no case does the work drip solution drainage into previous process tanks in the cycle. A straight-line cycle could provide the same sequence, but the load and unload stations would be widely separated and require additional load–unload equipment or additional personnel if manually loaded and unloaded. Another style tries to place process tanks such that those most sensitive to drippage problems and cross-contamination are placed in parts of the in-line layout where they would be least troublesome. The transport system is then programmed to take the work in a nonsequential order, skipping over stations and backtracking in order to complete the plating cycle. The carrying of work over other bars tends to leave salt buildup on the superstructure of the work bars, and potential contamination of any and all tanks in the line depending on which station the salt buildup eventually drops into.

Fundamentals of Electroplating

The essential components of an electroplating process are an electrode to be plated (the cathode); a second electrode to complete the circuit (the anode); an electrolyte containing the metal ions to be deposited; and a d-c power source. The electrodes are immersed in the electrolyte such that the anode is connected to the positive leg of the power supply and the cathode to the negative. As the current is increased from zero, a minimum point is reached where metal plating begins to take place on the cathode. The physics of this process has been the topic of many studies, and several theories have been proposed. A discussion of these theories can be found elsewhere (19).

Plate Thickness. In plating processes, plate thickness can be predicted knowing the cathode efficiency of a particular plating solution, the current density, and time of plating.

Plate thickness is an important factor in electroplating, in terms of both performance and economics. Corrosion resistance, porosity, wear, appearance, and several other properties are proportional to plate thickness. Minimum plate thicknesses are, or should be, specified as should the location, or check-point, where the thickness is to be measured. In some applications, such as threaded fasteners, maximum thicknesses should be specified. Root diameters of finer machine threads can be adversely affected by as little as 10 μm of plating.

A problem that affects the accuracy of the prediction of plating thickness is in estimating the actual current density. Current is not evenly distributed over the surface of the part being plated, rather, it takes the path of least resistance. Current also concentrates on sharper points, corners, and edges; even the shape of the plating tank can have an influence on the current distribution. The difference in current and, subsequently, the plate thickness distribution, is minimal when geometrically conforming anodes are part of the system, but this condition is not often achieved.

Cathode Efficiency. Faraday's law relates the passage of current to the amount of a particular metal being deposited; ie, 96,485 coulombs, equal to one Faraday, deposits one gram-equivalent weight of a metal at 100% efficiency. The cathode efficiency, an important factor in commercial electroplating, is the ratio of the actual amount of metal deposited to that theoretically calculated multiplied by 100%.

Throwing Power. It has long been observed that plate thickness distribution does not always follow the primary current distribution. Throwing power, a term used to describe the relative plate thickness distribution, was initially defined as the improvement in percentage of the metal distribution ratio above the primary current distribution ratio. A test method for the measurement and mathematical expression of throwing power, called the Haring Cell test, is available (20). Plating solutions vary widely in ability to exhibit good throwing power. As current is increased in a given plating process, a point is reached where the metal ion being deposited is not replaced in the solution film nearest the surface fast enough and a concentration polarization occurs, shifting some of the current to the unpolarized, lower current density areas. The effect is to reduce the plating rate in these higher current areas.

The values obtained using the Haring Cell test are only relative, but can be useful for comparison purposes. Copper cyanide plating solutions under a given set of operating conditions can give a throwing power value of 60%. On a particularly shaped part, the ratio of copper plate thickness was found to be less than 2:1, highest-to-lowest plate thickness. The same part, plated in a chromium plating solution having a 100% throwing power, had a thickness ratio of more than 20:1, and in an acid copper plating solution having about a 5–10% throwing power, had a thickness ratio of about 7:1.

A plating solution that exhibits a decrease in cathode efficiency with increasing current density exhibits better throwing power, ie, more even plate thickness distribution, than other solutions. Generally, the solutions having positive throwing power are those that contain some complexing agent for the metal ion being deposited. Complexing agents impede the transfer and reduction of the metal ion in the cathode film, promoting concentration polarization (see CHELATING AGENTS). Conversely, poorer throwing power is obtained from those plating solutions containing simple uncomplexed metal ion salts.

Examples of plating solutions having good throwing power include cyanide plating baths such as copper, zinc, cadmium, silver, and gold, and noncyanide alkaline zinc baths. Examples of poorer throwing power baths are acid baths such as copper, nickel, zinc, and hexavalent chromium.

Covering Power. The term covering power is often but erroneously used interchangeably with throwing power. These are two distinctively different terms (1), and a plating solution can have good covering power but poor throwing power. This is more often seen in barrel plating processes where the parts cover well but show wide thickness variations. Covering power refers to the lowest current density area where plating appears on a part. There are qualitative and semiquantitative tests for covering power, but no standard method is described by ASTM. One method simulates plating a tubular part by rolling a metal foil into a tubular shape, plating the rolled foil, unrolling, and measuring the distance that the plating covered on the inside of the foil.

Current Density. For a given electroplating process, as the current density is increased on the work, the point where the deposit becomes rough, coarse-grained, and takes on what is termed a burned and generally unacceptable appearance is eventually reached. The range from the minimum covering power to just below this burn producing current is the usable current density range. In electroplating, this range is not the same for all metal plating baths, and is influenced by metal ion concentration, chemical compositions, temperature, agitation, and anode-to-cathode spacing, as well as by the shape of the work and how it is positioned. In production plants, it is usually preferred to plate at the highest possible current density to shorten production time. This is not always done, however, especially when using lower throwing power plating baths, because increasing current decreases throwing power even further. For parts that require more even deposit thicknesses, lower current densities and longer plating times are used.

The Plating Tank. Materials of construction for the plating tank and the required auxiliary equipment are chosen either to be totally inert to the plating solution, or lined with inert material to protect the tank. In all cases, it is preferable to have a nonconductive inner surface of the tank to provide electrical as well as chemical resistance. A more recent requirement in many states is a containment surrounding the plating tank area to prevent accidental chemical spills from entering the environment (see TANKS). For alkaline plating solutions, mild steel materials are used; in acid plating solutions, other materials are used, depending on the chemical composition of the plating bath. Titanium and various stainless steel alloys, polytetrafluoroethylene (PTFE), Karbate, Hastalloys, zirconium alloys, and others are among the choices.

Sizing. There is no limit to plating tank size. Commercial electroplating lines have varied from liter-size tanks in benchtop wire plating lines to those holding 100,000 L and more. Tanks of over 250,000 L were used by the automobile industry when large amounts of plated trim were used. Tank width should allow for adequate anode-to-cathode spacing. In strip and wire lines, where the work is of a simple shape, anode-to-cathode spacing can be on the order of several millimeters to keep voltage requirements low. For plating work on racks, consideration should be given to the size and shape of the work to be plated. More complicated parts having deeply recessed areas require greater anode-to-cathode spacing.

In the simplest tank arrangement, there is a single work lane in the center and two anode bars equidistant from the center. If only one side of a part were to be plated, there would be no need for the second anode bar. A double-lane tank arrangement has two work lanes having an anode bar between and anode bars on the outside. Each anode is equidistant to the center of each work lane. Plating lines having more than double-lane plating tanks have been used, but these are much less common. The exception is those for wire lines. No standard rules appear to be followed in sizing tanks for rack plating parts having deeply recessed areas. Most of the time the part is made to fit whatever size tank is available. Fewer control and quality problems arise using anode-to-cathode spacing of 40 cm, rather than 15–20 cm, especially when plating parts with deeply recessed areas.

Plating tank depth requirements have changed. Plating tanks were historically much deeper than the length of the plating racks used, allowing for a substantial settling zone for particles that entered the plating solution from whatever source. Deeper tanks also allow room for installing pipe lines for filter headers, solution agitation, or air lines for air agitation. Rack plating tanks having a free-board of at least 20 cm and 30–60 cm deeper than the longest rack would reach, give a less troublesome plating tank. The length of the plating tank is related to production requirements, and whether the tank is part of a conveyORIZED automatic system.

Busbars. Fitting the tank for d-c power is usually accomplished using round copper busbars, both for supporting the anodes and the work or cathodes. Size of the copper bus is determined by the amount of current flow expected; 1000 amperes requires about 6.5 cm² of cross-sectional area. The bus is insulated from the tank, and any other sources of grounding, and connected to the d-c power supply. Shorter distances from the tank as well as fewer electrical connections keeps the voltage drop to a minimum.

Filters. Continuously filtering plating solutions are becoming much more common, except for hexavalent chromium plating solutions. Baths are filtered to remove any fine particulate matter that could codeposit with the electroplate and cause roughness. Inlets to the filter(s) are usually located near the bottom of the plating tank, often in several areas or through a perforated header-pipe. The discharge is best placed near the top of the solution and directed horizontally. The filter flow is a function of retention size among other factors. Depending on the kind of plating bath and quality requirements, the filter retention size may vary from about 25 µm down to 0.2 µm. Platers refer to filter capacity as the tank turnovers per hour; the media retention should be included. A filter turning over five-tank volumes per hour through a 15-µm media would do less than one turnover through a 1-µm media. Filter media has to be chosen carefully to avoid introducing any soluble impurities and to avoid dissolving any of the media itself.

Heating and Cooling. In some plating tanks heating or cooling is required preferably using a means of automatic control. Heating coils, plate coils, and other forms of immersion heaters are placed along a tank wall, preferably in an area that has good solution agitation. Heating can also be accomplished using external heat exchangers where the solution is pumped from the plating tank. In-tank heaters have to be properly located and insulated to avoid becoming part of the plating electrical circuit. Smaller tanks are often heated with quartz electric heaters. Steam (qv) coils and steam-heated external heat exchangers are used in large installations (see HEAT EXCHANGE TECHNOLOGY).

Cooling is required in some plating solutions. High currents in relatively smaller solution volumes often bring about rising temperatures and solutions which can require cooling to continue, even if heating was required to begin plating. Other solutions require cooler than ambient temperature to perform properly. Cooling can be accomplished using cooling water coils, if demand is not high, or refrigeration units, commonly called chillers (see REFRIGERATION AND REFRIGERANTS).

Temperature control is important to all electroplating baths where good quality is required, because many properties are affected with only a few degrees of temperature change. Although a given plating bath may operate over a wide temperature range, it most often has a much narrower range where the desired properties are maximized. It is common practice to specify room temperature as a plating solution's operating parameter. In most cases this is too broad a range for good consistent quality. The preferred range is 5°C or less from a set temperature. An automatic temperature control is a considerable benefit to quality plating.

Anodes. There are two types of anodes: soluble and insoluble. Most electroplating baths use one or the other specifically; however, a few baths use either or both. Chromic acid plating baths use insoluble anodes; alkaline zinc cyanide baths use both; noncyanide alkaline zincs may use either. Soluble anodes are designed to dissolve efficiently with current flow and preferably, not to dissolve when the system is idle. A plating solution having the anode efficiency close to the cathode efficiency provides a balanced process that has fewer control problems and is less costly. If the anode efficiency is much greater than the cathode efficiency and there are only small solution losses, the dissolved metal concentration rises until at some time the bath has to be diluted back or the excess metal has to be reduced by some other means. If the anode efficiency is less than the cathode efficiency, the dissolved metal decreases, pH decreases, and eventually metal salt additions and other solution corrections are required. Based on the cost of metal, it is usually considerably more economical to plate from the anode rather than add metal salt. Copper cyanide, for example, costs about twice as much to add than to dissolve a comparable amount of copper anode. Additionally, the anion added with the metal salt may build up in the plating solution.

Faraday's Law applies to the anode as well as to the cathode; ie, the total reaction at the anode is proportional to the current, and much like the cathode, the anode efficiency varies with the current density. As the current on the anode is increased, the anode efficiency decreases, slightly at first, until it reaches a point at which the anode metal cannot dissolve fast enough through the anode film. The first stage of dissolution for the soluble anode is the oxidation of the metal followed by dissolution of the oxide. When the oxide dissolution rate is less than the oxidation rate, polarization of the anode takes place. The oxide film builds up in sufficient thickness to form an insulating coating, and the current decreases rapidly. The thick anode films can dislodge at

the reduced current and remain as particulate matter until they slowly redissolve. As particulate matter, the dislodged anode films form a significant source of roughness because they can co-deposit with the metal on the work before redissolving. While the anode is polarized, oxygen and other gasses can be given off from the anode. Solutions containing chlorides or bromides can emit chlorine or bromine, which are hazardous. Anode current densities should be maintained well below the point at which polarization occurs. This can be accomplished by increasing the anode area and lowering the current.

Agitation of the solution around the anodes and some chemical variations can increase the anode current density range before polarization. Lower pH and chloride ion have been used to improve anode corrosion in acid nickel baths. Higher free cyanide, Rochelle salts (sodium potassium tartrate), and proprietary substitutes improve anode efficiency in cyanide copper solutions. Higher alkalinity improves zinc anode corrosion in the alkaline zinc solutions; lower pH increases zinc anode corrosion in acidic chloride zinc solutions. Some platers use anode bags to prevent particulate anode and anode film material from reaching the work. However, the solution flow through the bag material may be inadequate and add to the polarization problem. The weave opening of the bag material is important to maximize solution flow without passing excessive particulates. Anode bags must be replaced once plugged.

Soluble anode materials are not always a pure metal. In acid, low chloride nickel solutions, pure nickel does not corrode well, and small amounts of specific impurities are added to make the nickel more active, allowing more efficient dissolution. For example, since the early 1960s, nickel anode material containing a small amount of nickel sulfide [16812-54-7], NiS, has been commercially available and important in nickel sulfamate [13770-85-3], Ni(H₂NO₃S)₂, plating baths. These anodes corrode at a lower potential than pure nickel or other nickel anode materials (see NICKEL AND NICKEL ALLOYS).

Copper anodes for use in acid copper plating solutions preferably contain a small amount of phosphorus [7723-14-0], usually 0.03–0.04 wt % , which retards chemical dissolution of the copper and thus the subsequent copper build-up. Typically, acid copper plating solutions increase in copper and require periodic dilution. Additionally, additives for brightening acid copper baths tend to last longer in plating tanks using phosphorized copper anodes. In cyanide copper solutions, phosphorized copper anodes should not be used.

Since the 1960s titanium mesh anode baskets have been used (21), especially in nickel plating solutions. Nickel anodes in the form of small round buttons and pellets combined with the titanium anode basket allows a constant anode area to be maintained with a minimum of effort.

Insoluble anodes are used exclusively in some plating baths. Chromium plating solutions utilize lead–tin, lead–antimony, or lead anodes. Gold and other precious metal plating processes use stainless steel anodes, keeping inventory costs down. Other plating processes make use of insoluble anodes, on occasion, either in an effort to reduce metal ion buildup or as an auxiliary anode to direct the current into recessed or low current density areas. Platinum plated titanium and platinum plated niobium insoluble auxiliary anodes are often used in decorative plating lines for plating complex-shaped parts.

Whenever insoluble anodes are used, the pH of the plating solution decreases along with the metal ion concentration. In some plating baths, a portion of the anodes is replaced with insoluble anodes in order to prevent metal ion buildup or to reduce metal ion concentration. Lead anodes have been used in acid copper sulfate baths, and steel anodes have been used in alkaline plating baths.

The use of insoluble anodes can also result in side effects. In alkaline cyanide solutions, the generation and buildup of carbonates is accelerated remarkably, along with a significant reduction in alkalinity. In acid solutions the pH decreases as well, requiring frequent adjustments. In sulfamate nickel plating solutions, insoluble anodes, and even slightly passive soluble anodes, partially oxidize the sulfamate ion to form sulfur-bearing compounds which change the character and performance of the deposit (22).

d-c Power Supplies. The Daniell cells and batteries of the mid-nineteenth century, replaced by motor-generator sets in the first half of the twentieth century, have given way to solid-state silicon rectifiers. A variety of modifications in power supplies has become available. Stepless controls, constant current, and constant voltage are popular. Automatic current manipulations are also getting more use. Current interruption cycles where the current is periodically discontinued have been used in alkaline baths. Cycles on the order of 10 seconds on to 2 seconds off have been employed (15) in cyanide copper baths using a variety of time and current density cycles. One system uses about 45–60 s at 500 A m² and 15–20 s reverse polarity at about 200 A m². The newer forms of current manipulation are called pulse cycles because the cycle is typically on the order of milliseconds. A number of variations are possible (23).

The d-c output from a rectifier can have a portion of modulation from the a-c input, called ripple. Modern rectifiers usually have 5–10% or less ripple at full load. High ripple can have some bad effects, especially in chromium plating where ripple can cause dullness and poor coverage. The exact amount of ripple to cause the problem varies with the chemical composition, and a finite figure is difficult to set. However, there is usually some effect when ripple exceeds about 20–30%. Some dullness has been encountered using alkaline noncyanide zinc plating baths, but few of the other plating solutions have been reported as being sensitive to ripple. In all cases, the d-c power supply needs a means of control, stepless preferred, and appropriate ammeter and voltmeter. Ampere-hour recorders are convenient for making additions and keeping records.

Preparation for Plating

Preparation cycles vary depending on the particular substrate being electroplated. High carbon steels, low carbon steels, free-machining steels (lead or bismuth), HSLA steels, stainless steels, cast-iron, and high nickel alloys all require special considerations in preplate treatments, as do the many nonferrous alloys and various plastic substrates. Often only slight variations in substrate composition significantly influence the preparation process. Heat treating variations also contribute to the complications in preparation. Determining an optimum preparation process for a given material often becomes a matter of trial and error. Some guidelines can be found in the literature (24,25) and from suppliers of proprietary preparation products. Several ASTM guides are

ASTM guide no.	Title
B253	<i>Preparation of Aluminum Alloys for Electroplating</i>
B322	<i>Cleaning Metals Prior to Electroplating</i>
	<i>Practices for preparation of and electroplating on:</i>
B630	<i>Chromium (Electrodeposits) on Chromium</i>
B281	<i>Copper and Copper-Base Alloys</i>
B320	<i>Iron Castings</i>
B319	<i>Lead and Lead Alloys</i>
B480	<i>Magnesium and Magnesium Alloys</i>
B629	<i>Molybdenum and Molybdenum Alloys</i>
B558	<i>Nickel Alloys</i>
B343	<i>Nickel (Electrodeposits) with Nickel</i>
B727	<i>Plastic Materials</i>
B242	<i>Steel, High Carbon</i>
B183	<i>Steel, Low Carbon</i>
B254	<i>Steel, Stainless</i>
B481	<i>Titanium and Titanium Alloys</i>
B482	<i>Tungsten and Tungsten Alloys</i>
B252	<i>Zinc Alloy Die Castings</i>

Poor preparation of the substrate can result in loss of adhesion, pitting, roughness, lower corrosion resistance, smears, and stains. Because electroplating takes place at the exact molecular surface of a work, it is important that the substrate surface be absolutely clean and receptive to the plating. In the effort to get the substrate into this condition, several separate steps may be required, and it is in these cleaning steps that most of the problems associated with plating arise.

Cleaning Methods. The soils encountered in electroplating processes can be organic, eg, oils, greases, and other cleaning compounds, and

also inorganic, such as oxides and heat-treat scales. Some plating baths can clean surfaces and thus tolerate minimally cleaned surfaces; others need surfaces cleaned to near perfection. Chromium plating from chromic acid solutions is a good example of the former. In practice, some hard-chromium platers depend on the strong oxidizing nature of the chromic acid to remove residual oils. A short period of reverse current before plating provides an etched surface, often clean enough to plate. Another plating bath that possesses good cleaning properties is the high cyanide zinc plating solution. Historically little or no cleaning was used when barrel-plating in high cyanide zinc (or cadmium) plating solutions. Work would be tumbled without current in the plating solution until clean enough to plate. Quality, however, was not always guaranteed. Use of high cyanide zinc solutions has been in decline since the late 1960s primarily because of environmental regulations and the high cost of waste management.

The majority of plating baths require excellent cleaning. Substrate surfaces have to be free of even the slightest of soil-containing films or oxides as the work enters the plating solution. No simple, universal cleaning cycle exists for electroplating. Several methods and cleaning solutions may be used in a single-plating process (26,27).

Mechanical Cleaning. Parts should go through an alkaline cleaning step to remove organic soils before mechanical cleaning, which employs some form of abrasion. Dry blasting uses sand, aluminum oxide or silicon carbide grits, glass beads, or similar abrasives (qv). Wet blasting, vapor honing, wet tumbling with abrasive media, and vibratory finishing are also used. These methods are often required to remove heavier heat-treat scales. Alternatively, a strong pickling acid treatment, which has the severe disadvantage of promoting hydrogen embrittlement especially in hardened steels, must be used.

Solvent Cleaning and Vapor Degreasing. Heavy organic soils of many types including waxes are removed by solvent cleaning and vapor degreasing. The latter also results in dry parts, but because solvents such as the chlorinated hydrocarbons and chlorofluorocarbons (CFCs) are employed, usage has declined and is expected to continue to decline (see CHLOROCARBONS AND CHLOROHYDROCARBONS). In many cases, aqueous cleaning is being substituted. Vapor degreasing equipment for preventing loss of the volatile solvents has been undergoing improvement, however, in an effort to allow continued use. The 1990 Clean Air Act in the United States has set a timetable for unilaterally eradicating such materials as the CFCs and the chlorinated hydrocarbons, eg, methyl chloroform [71-55-6] (1,1,1-trichloroethane). In 1989, 50,000 vapor degreaser installations in the United States used about 260,000 metric tons of chlorinated solvents of which 60% was methyl chloroform (28). The number of degreasers used only in electroplating is not known, but most plating plants had at least one vapor degreaser or solvent cleaning tank available.

Diphase Cleaning. Use of an organic solvent top layer over a mild aqueous cleaner solution was popular for aiding in the removal of buffing compounds, especially on zinc base die castings. Use of this diphase cleaning has declined with the decline in buffed zinc castings.

Spray Cleaning. The advantages of spray cleaning are high agitation and physical force. The limitations are in focusing the sprays to hit all surfaces and in having to limit spray composition to nonfoaming constituents. Spray cleaning use has also declined because of the decline in plating of buffed substrates.

Ultrasonic Cleaning. Both organic solvents and aqueous cleaners can be used in ultrasonic cleaning tanks. Ultrasonic rinsing has been helpful in processing porous substrates. Although the chemical costs can be low, equipment costs are not. The use of ultrasonics has grown, but not significantly.

Soak Cleaning. A common method in plating plants is soak cleaning, which is the first immersion cleaning step on the plating line. Soak cleaners can be characterized as alkaline or acid. Alkaline is the most common. Formulations for alkaline soak cleaners can contain some or all of the following: sodium hydroxide, sodium carbonate, sodium silicates, trisodium phosphate, tetrasodium pyrophosphate, borax, sodium gluconate or sodium heptonate, and minor amounts of chelating agents (qv) such as EDTA or NTA. In addition, formulations can contain soaps, although these are becoming less common, and other surface-active compounds. Surface-active components may be detergents, wetting agents, or surfactants (qv). Other minor constituents may include defoamers (qv), dispersants (qv), which prevent redeposition of soil; and materials that aid in rinsing, solubilizing, or inhibiting attack on the substrate. Some formulations substitute potassium salts for the sodium for increased solubility at an increase in cost. Whereas typical formulations can be found in the literature, most plating plants utilize proprietary cleaning compounds formulated by suppliers specializing in cleaning compounds for the plating industry.

Soak cleaner formulations have been modified over the years. Older formulas contained higher caustic contents and required operation at high concentrations and high temperatures, often at a rolling boil. For some stubborn soils such as waxes (qv), this type of cleaner still finds use. Generally, the trend has been toward greater use of surfactants, lower temperature operation, and lower caustic contents. Most of the time, the plating plant has no control over the type and quantity of soil on the work received. These soils are usually neither single compounds, nor easily identifiable. Besides dirt, typical soils are metal-working compounds such as rust preventatives, forming compounds, cutting fluids, die lubricants, and stamping oils. Soils introduced from improved metal-working often require new cleaner formulations.

Cleaning is accomplished by one or more of several mechanisms, eg, saponification, solubilization, emulsification, wetting, sequestration, and deflocculation. In soak cleaner formulations, there is usually both an inorganic portion and an organic one. Optimum concentrations and operating temperatures exist for the various cleaner blends in a working bath. Higher temperatures have an expected benefit in removing soils because of increased activity of the components and increased reaction rates. The maximum performance of a cleaner, however, is not necessarily at the highest temperature. Surfactants are typically less soluble in solutions having higher salt contents and at higher temperatures. When the solubility limit of the surfactant is exceeded, some of the surfactant separates and the solution becomes turbid. Under these conditions, the separated surfactant agglomerates and usually collects in oily appearing masses on the surface of the cleaner tank. At the first sign of turbidity, the cleaner is said to behave almost as a diphase cleaner, the surfactant acting more like an oil-loving solvent. It is at just below this cloud point that most cleaners perform best and most economically.

The Water-Break Test. It is the nature of surfactants to be substantive, ie, have high attraction to solid surfaces. Additionally, oils are very substantive. Thus, concentration of surfactant on the surface of the work is greater than in the body of the solution. The surfactant is also tenacious and difficult to rinse. In the course of removing organic, oil-like soils from the work surface, some of the soil is solubilized by the surfactants, but the resultant oil-surfactant mixture still retains some of the substantive properties of the surfactant, ie, it has become water-soluble, but clings. A common visual test to determine cleanliness of the work is to observe how the surface rinses; if the surfaces drain evenly when rinsed using clean water, holding a sheet of water, it is said to be water-break-free and clean. If rinse water beads up in droplets on the part surface or does not hold a sheet of water as it drains, it is said to water-break and is in need of additional cleaning. However, if this test is run on surfaces having surfactant-solubilized oil films, the water-break test is not definitive. Thus, the part has to be thoroughly rinsed, acid dipped in clean acid, and thoroughly rinsed again before using the water-break test. Neither does the water-break test reliably detect inorganic films. Thus, the water-break test indicates when a part is dirty, but not necessarily when it is clean.

Controlling Soak Cleaners. The relationship of an optimum temperature-concentration value for good cleaning requires better temperature control and more frequent analysis and adjustments than is the rule in many plating plants. Larger soak cleaner tanks are typically heated with internal steam coils; smaller tanks may be electrically heated. Direct-fired tanks are becoming much less common. Heating a solution is much more efficient if the solution is agitated, avoiding localized overheating, temperature stratification, and resulting in improved heat-transfer. External heat exchangers having adequate pumping rates are much preferred over still tanks. An added benefit is the improved cleaning action with agitation.

Automatic temperature controllers are recommended, and should be capable of holding a set temperature within a range of 4°C or less. Concentrations of cleaning solutions can be controlled through simple titrations. Many suppliers of proprietary cleaners provide simple test kits that can be used at tankside by unskilled help. Frequency of analyses depends on the usage of the cleaner tank and the drag-out. It should be frequent enough so that required additions should be less than 10% of the set concentration. This could occur once per day with high usage, or once per week at lower use. Additions should be kept reasonably small and spaced out over a reasonable period to avoid sudden changes. When using dry, powdered, or crystalline cleaning compounds, additions to hot cleaner tanks can react violently, spewing hot alkali out of the tank. It is safer to make additions at the end of a shift, when the solution can be cooled and additions sprinkled in slowly using good agitation. The use of liquid cleaner concentrates eliminates this hazard, and probably accounts for much of the increased interest in liquids.

Improvements. Soak cleaner tanks equipped with an overflow weir and a sump, plumbed through the bottom of the sump to a recirculating pump provide several benefits, including extra agitation of the solution; aid in heat transfer; minimal temperature stratification throughout the tank; a constant level; aid in keeping a more constant cleaner concentration; an area for collecting and removing separated oils (in the sump) and an area to install an oil-separating device; and an area for making cleaning compound additions.

Further, more attention is being given to the oil content of cleaner wastes. This attention is expected to increase substantially in the near future as tighter environmental regulations come into play.

Cleaners. Cleaner formulations can be classified with respect to the ability to emulsify oils, keep them in solution; or to reject oils, split them out of solution. The latter type loosens the oil, and the separated oil does not dissolve but is rejected, and usually floats to the top of the cleaner solution. For the cleaner that rejects oil, the tank needs to be skimmed constantly. This is most easily accomplished using the overflow weir, sump, and recirculating pump arrangement. Oil-rejecting cleaners last longer.

When using emulsifying soak cleaners, a limit is reached in use where the oil-holding capacity is exceeded. This capacity varies with the formulation, but eventually the contaminated cleaner begins to act as a soil. Some emulsifying formulations are capable of rejecting some oil after reaching the limit on emulsifying, and properly skimmed, some of these cleaner tanks have lasted for years. Purification of oil-laden cleaners has been carried out successfully in production tests. Comparative tests run in the 1950s using periodic activated carbon treatment on one cleaner line resulted in increased cleaner tank life on that line.

Acid Soak Cleaners. Benefits in processing steel using acid soak cleaners include the ability to both clean and derust in the same tank. Acid soak cleaners, which are less popular, more expensive, and more corrosive than alkaline cleaners, have found use in some lines when plating on aluminum alloys and copper alloys. Formulations for acidic soak cleaners have not shown high soil-load capacity, and are used for relatively clean stock. More common formulas are based on phosphoric acid mixtures with surfactants and organic emulsifying agents. Acid soak cleaners for aluminum may contain nitric acid mixtures. Excellent rinsing is required after these to prevent any carryover into sensitive plating solutions.

Removing the soak cleaner film from the work, whether it is alkaline or acid, takes a good portion of the preplate process cycle. If the soak cleaner has performed well, the organic soils have been solubilized completely, and the water-insoluble soil film originally on the work has been replaced with a surface-active but water-soluble film. This film now contains the elements of the cleaner along with the cleaner-soil constituents. Rinsing in clean water is essential to reducing this tenacious film.

Electrocleaning. Application of a low voltage d-c current through a conductive alkaline cleaning solution using the work as one electrode and usually steel suspended from busbars on either side of the work as the other, is called electrocleaning. The equipment and process has many of the elements of the electroplating process. When the work is connected to the positive pole of the rectifier, the process is referred to as anodic or reverse cleaning, ie, it is reverse to electroplating. Oxygen generated in copious quantities on the work provides agitation at the immediate surface of the soil-part interface. Anodic cleaning is used primarily on steel. Special precautions and milder conditions are used when cleaning zinc-based die castings and copper or copper alloys.

Cathodic or direct cleaning uses the work as the cathode; hydrogen is generated at the work surface. The reducing environment on the part often helps in loosening stubborn scales. Direct current prevents attack of the substrate, but metallic impurities in the cleaning solution may deposit as nonadherent smuts. Direct cleaning is used in the preparatory process for many difficult-to-plate metal substrates. Copper and copper alloys do not tarnish with cathodic cleaning. Thus direct cleaning is sometimes used for most of the time cycle, then the polarity switched to anodic cleaning for the last 30 seconds.

Cathodic cleaning of hardened steels can lead to excessive hydrogen embrittlement. Usually, cathodic cleaning has been avoided on steels having a Rockwell hardness (HRC) greater than about 40 HRC; more recently, this has been lowered to 35 HRC. Hydrogen embrittlement can be relieved by baking for a minimum of 3 hours at $190 \pm 15^\circ\text{C}$ (29). This baking process is required after plating hardened steel regardless of how the part was cleaned, but it is considered good practice to minimize the introduction of hydrogen wherever possible. The ASTM guides, such as those for engineering chromium (30) and autocatalytic nickel-phosphorus (31), give explicit baking times and temperatures. Specifications applicable to each particular process should be followed.

Surfactants in Electrocleaners. Surfactants typically consist of a long-chain hydrocarbon molecule having a solubilizing or water-loving group which can be anionic, cationic, or nonionic when solubilized. Thousands of surfactant products are marketed, usually under trade names (32). In commercially formulated electrocleaners, surfactants are usually anionic, and often mixtures of anionics and nonionics.

Many plating plants use electrocleaners both anodically and cathodically through the use of reversing switches or in two separate tanks. Not all anionic surfactants do well in cathodic cleaning. Proprietary electrocleaners are commonly used and highly recommended.

Operation and Function. Overheating electrocleaner solutions can result in surfactant loss, as for soak cleaners. Tank construction and electrodes can be steel. Current distribution is easier to optimize and control when the inside of the electrocleaner tank is lined with an inert, nonconductive lining. An overflow weir, sump, and recirculating pump are recommended. When plating barrels are used electrocleaning can take 10 minutes or more compared to 1 minute on a plating rack. Higher voltages are also needed.

When an electrocleaner follows a soak cleaning step, its main function is to remove the soak cleaner film left on the part. If the soak cleaner has done its job, the soil remaining on the part has been made more soluble. There are some significant differences in the chemical compositions of electrocleaners and soak cleaners. Electrocleaners are not usually designed to remove heavy organic soils, and are formulated using much lower concentrations of surfactants. Excess surfactants should be avoided, because the more surfactant, the harder it is to rinse. The goal is to get the cleaned part to the plating tank without any residual wetting agent films.

A secondary function of the electrocleaner is to aid in removing light smuts or scale. Complexing agents may assist in this function, but the possible effect on the waste treatment process has to be considered. Cyanide, an effective smut remover, is no longer used. Heavier carbon smut, especially that from etching of steel substrates, is difficult to remove. This problem is more pronounced in high carbon steels. It is preferable to minimize etching.

Electrocleaners are used cathodically to activate normally passive substrates such as nickel, nickel alloys, and stainless steels. When cleaned anodically, these materials become much more difficult to activate. Unless fully activated, little or no adhesion is obtained. Anodic cleaning dissolves chromium and is commonly used for stripping chromium deposits. The presence of dissolved chromium in the electrocleaner is undesirable and causes uneven etching, leaving patterns that are still visible after plating. Sodium hydrosulfite in small quantities is sometimes used to reduce the chromium. Proprietary compounds based on reducing sugars are safer. Chromium is a common impurity in nickel-chromium decorative plating lines because the plating rack carries the chromium, either as the solution or as chromium-plated contact tips or both, back through the cleaning cycle. Less well-known but capable of causing problems, stainless steels can be anodically attacked, albeit slowly, in strong alkaline solutions and chromium dissolved from the alloy surface. Stainless steel anodes in alkaline solutions are sometimes, erroneously recommended; mild steel is usually adequate.

Alkaline Derusting. Adding highly alkaline solutions that can dissolve light rust is a form of electrocleaning dealing with inorganic soils on steel. Applying a d-c current increases this capability to a practical degree. Further improvements in the rate of removal are possible with the addition of certain complexing agents and the use of periodic current reversal cycles. A typical formulation for an alkaline deruster is about 80% caustic soda and 20% sodium gluconate of which 100–250 g/L are used. Elevated (65–90°C) temperatures aid conductivity and reaction rates. Current density should be 500 A/m² or more with current cycles of about 60 seconds with the work anodic and 30 seconds cathodic. Work is removed on the anodic cycle. Time depends on several factors, and can be as long as several minutes. Some rust scales may require acid pickling to remove them completely.

Cleaning Cycles. Cleaning cycles usually start with a soak cleaner followed by electrocleaning. If work is heavily soiled, however, a precleaning step may be required. Buffing compound residues fall into this class. Solvents, diphasic mixes, power washers, and other presoaks are used. If instead work is received in a fairly clean state, the high surfactant soak cleaner is omitted from the cleaning cycle. The electrocleaner formulation has less tendency toward leaving substantive films and is preferred. In some plating lines, there are more problems removing the soil-remover than in removing the soil.

Combination soak-and-electrocleaner products are touted for use when normal soak cleaner is too much and a normal electrocleaner is too little. These combination cleaners are often electrocleaners with additional surfactants. The results are often hard-to-rinse electrocleaner or a low capacity soak. The term heavy-duty, often used to describe cleaners, can refer to high caustic content or to good soil-removing, high soil-load capacity.

Rinsing. Rinsing is extremely important to the overall operation of a plating line. Work traveling from a processing solution carries a film on its surface from that solution. This amount of drag-out varies with the shape and surface area of the part, along with the nature of the solution. Drag-out can reach considerable quantities, eg, from 20–80 mL/m² (33).

To reduce costs of the more expensive plating solutions and decrease the amount of hazardous or regulated material in a waste stream, recovery and reuse of the drag-out is a common practice. This is done by closing off the water flow to the first rinse tank following the process tank, and periodically

returning the accumulated solution to the process tank. Used after nickel plating, a single saver-rinse can recover 60° of the nickel dragged out. This increases to 80° if two saver-rinses are used. In precious metal plating lines, there are several non-flowing rinse tanks (saver-rinses) following the plating tank. In chromium plating, the drag-out is processed through an evaporator before returning it to the plating tank.

Recycling drag-out losses can have the unwanted effect of concentrating impurities in the plating tank. Thus, it is good practice to purify the drag-out before returning it to the process tank. Filtration (qv) through activated carbon is helpful on most nonchromium solutions; cation exchange has been useful with chromic acid drag-out solutions, before it is evaporated. The impurity concentrating potential makes it even more important to use effective prerinsing with good quality water, ie, before work enters the process tank. Each processing solution is a serious contaminant to the next.

Good rinsing requires good water, not too cold, vigorous agitation, and time. Many seasonal plating problems, especially in northern climates, can be traced to colder winter rinses. Water at 5–7°C is a poor rinse; water at 30–35°C gives a good rinse. Time and agitation allow the rinse water to penetrate, to dilute, and remove the often substantive films. A two-minute dip in each agitated rinse has often produced good work having good adhesion, when one minute dips failed.

Dual- or triple-rinse tanks follow most process solutions. Additionally, many lines utilize counterflow or cascade rinsing with the multiple tanks plumbed so freshwater flows from one tank to the next counter to the direction of the work. This method of conserving water is recommended procedure for all plating lines. A double counterflow rinse of 70 L h could give equivalent rinsing to 2250 L h using a single rinse in reducing a concentrated solution of 2200 g L to 7.5 g L. For a triple counterflow rinse, the rate dropped to 30 L h (34).

An example of rinse stages in a relatively simple plating cycle for plating an engineering nickel on a common, low carbon steel having varying soil conditions is shown in Table 1. On the longer suggested cycle for heavily soiled, rusted stock, 13 of the 19 stations are some form of rinse.

Table 1. Processing Cycle for Engineering Nickel on Low Carbon Steel^a

Station	Heavy soil rust and scale	Heavy soil no rust or scale	Light soil no rust
1	soak clean ^b	soak clean ^b	
2	rinse	rinse	
3	electroclean	electroclean	electroclean
4	rinse	rinse	rinse
5	rinse	rinse	rinse
6	acid pickle ^c	acid dip ^c	acid dip ^c
7	rinse	rinse	rinse
8	rinse	rinse	rinse
9	electroclean ^d	^d	^d
10	rinse		
11	rinse		
12	acid dip, mild		
13	rinse		
14	rinse		
15	nickel plate	nickel plate	nickel plate
16	saver-rinse	saver-rinse	saver-rinse
17	rinse	rinse	rinse
18	rinse	rinse	rinse
19	hot rinse and dry	hot rinse and dry	hot rinse and dry

^a For high carbon (0.4° C and more), the desmutting steps 9–14 are required.
^b Preclean, option before soak.
^c For high carbon steel, keep mild as possible.
^d Alkaline desmut option.

Spray rinsing can be efficient because of the impingement; misting at the exit of a processing tank can reduce drag-out. Barrel plating lines need much more rinsing time and special consideration because of the relatively high drag-out rates.

Acid Dips and Acid Pickles. Acids are used to remove inorganic soils. A distinction is made between acid dips and acid pickles. Where there is no rust or scales, a dilute acid dip is used for activation and as a rinse aid. Caustic residues from cleaners are notoriously difficult to remove with water rinsing alone. By contrast, strong acid pickles are used to remove rust and scale.

The most commonly used acids in preplate processes are hydrochloric acid, HCl, more common for rust and scale removal, and sulfuric acid, H₂SO₄ (see HYDROGEN CHLORIDE; SULFURIC ACID AND SULFUR TRIOXIDE). Usually 30–60° by volume of 20° Bñ hydrochloric acid is used. At room temperature, hydrochloric acid removes complex iron oxide scales in a shorter time than sulfuric acid, produces a smoother, whiter surface, and holds more iron in solution. On the negative side, hydrochloric acid is more volatile and gives off corrosive fumes even at room temperatures. Ventilation is required. Hydrochloric acid pickles are operated at room temperature, always below 40°C. Sulfuric acid solutions increase in pickling rate with concentration up to about 25° strength; rates then decrease to essentially zero at full strength sulfuric acid. The pickling rate of 5° sulfuric acid is about 10 times faster at 60°C than at 18°C; at 10° sulfuric acid, this increases to 15 times faster (35).

Heat scales and rust on steel are not of even composition or thickness, and parts are subject to uneven pickling. To minimize over-etching, inhibitors are sometimes used to slow the attack on the clean steel. Many inhibitors produce strong adverse effects if carried into the plating solution. A second electrocleaning step is used after the pickle when inhibitors are employed.

Acid dips are more likely to be dilute sulfuric acid, 2–5° by volume of 66° Bñ sulfuric acid. The hazards of handling and diluting sulfuric acid have spurred use of proprietary dry acid salts, a common ingredient of which is sodium bisulfate. Proprietary compounds may contain wetting agents that require good rinsing to remove. Other acids have been used for specific substrates. Leaded steels and leaded brasses use pickles and acid dips of fluoboric acid (see FLUORINE COMPOUNDS, INORGANIC). Sulfuric acid should be avoided on leaded alloys. Free-machining steels having about 0.5° lead in the surface are treated with 10–20° by volume of 42 wt° fluoboric acid to avoid formation of insoluble lead salts. When strong pickling action is not required, citric acid [77-29-9] and proprietary citric-based products work well for leaded steel and copper alloys. Modified citric acid dips are being used successfully on steels such as chromium–molybdenum and nickel–chromium–molybdenum steels. Citric acid has been used in processing zinc-base die castings. Sulfamic acid, a strong acid which forms soluble lead compounds, is often mentioned for use on lead or lead-containing alloys. Sulfamic acid tends to bring out smut when used on high (>0.4%) carbon steels, and strong solutions decompose with time.

Acids containing fluoride, or mixtures of other acids and fluoride salts, are used on silicon-containing alloys. Sulfuric acid and ammonium acid fluoride mixtures are used on high silicon steels; nitric acid with ammonium acid fluoride is common in processing high silicon–aluminum alloys. Higher magnesium-containing aluminum alloys often are treated with dilute sulfuric acid; other aluminum alloys with nitric acid.

Chromic acid is seldom used in preplate acid dips, but has had special applications in other preplate treatments. Chromic acid pickles or prepickles are part of some processes for plating on certain magnesium alloys. Chromic acid-containing bright dips are sometimes part of the preplate treatment on copper alloys; and to a less extent on zinc-base die castings. In plating on plastic substrates, strong chromic acid solutions are used to etch the base for

providing adhesion. Whenever chromic acid is used in the preplate process, it is important that no residual hexavalent chromium remains on the surface. Only a very small amount of chromic acid can yield poor adhesion and blistering, or stop the plating completely.

Electropickling. At times, electrolytic pickling is used; steel can be treated anodically in strong sulfuric acid, 70° o. If the sulfuric acid is too dilute, anodic etching can be severe. In bright nickel plating of steel, pretreatment in a cathodic 5–10° o sulfuric acid has reduced subsequent pitting of the nickel. Cathodic treatment prevents etching of the steel.

Cathodic treatment of steel parts in acids could be expected to contribute significantly to hydrogen embrittlement of the part if the steel has been previously heat treated to over 40 HRC. Cold-worked steel is susceptible at a lower hardness. Some work shows more embrittlement from the plating bath than from preplate treatments (36).

When plating any substrate less noble than copper, only a few mg L of dissolved copper in the acid baths can adversely affect adhesion. Coatings can be too thin to be visible, yet contribute to poor adhesion. Small additions of thiourea have been used to prevent copper immersion, but it acts as a potent inhibitor, and work should be re-electrocleaned after the acid. Work should be exposed to the mildest acid treatments possible. Over-etching should be avoided.

Properties, Specifications, and Test Methods Standard test methods are required to measure the properties of electroplated materials. Documents on plating specifications for many phases of the plating process are published by such organizations as the Federal government, the military, ASTM, ISO, SAE, etc. An excellent cross-index of these is available (37).

Plate Thickness. Thickness of the plate should always be specified as should the locations on the work where the thickness is to be measured. Generally, thicker deposits perform better, but there are notable exceptions. Mating parts, eg, fasteners having fine machine threads, are not usable if over plated. Machine-threads are usually plated to 10 µm or less, depending on tolerances. Additionally, gold-plate over nickel does not solder well if too thick; thus, gold is usually 1–2 µm or less. Chromium, plated for decorative purposes from the conventional chromic acid bath, tends to macrocrack above about 0.7–1.0 µm.

Reference 38 is a good guide to the selection of plate thickness test methods. Test methods may vary with the purity and electrochemical activity of the deposit. Metals deposited from commercial plating solutions are seldom pure. For example, zinc deposits from the three commonly used baths, ie, cyanide, chloride, and zincate, vary significantly in purity and activity (39). Standard ASTM test methods for determining plate thickness are

ASTM guide no.	Title
B487	<i>Microscopical Examination of a Cross Section</i>
B499	<i>Magnetic Method for Nonmagnetic Coatings on Magnetic Basis Metals</i>
B504	<i>Coulometric Method</i>
B530	<i>Magnetic Method for Nickel Coatings on Magnetic and Non-magnetic Substrates</i>
B555	<i>Guide for Measurement of Electrodeposited Metallic Coatings by the Dropping Test</i>
B556	<i>Guide for Spot Test for Thin Chromium Coatings</i>
B567	<i>Beta Backscatter Method</i>
B568	<i>X-Ray Spectrometry</i>
B588	<i>Interference Microscope Technique</i>
B659	<i>Guide for Measuring Thickness of Metallic and Inorganic Coatings</i>
B681	<i>Light-Section Microscope</i>
B748	<i>Measurement of Cross Section with a Scanning Electron Microscope</i>
B767	<i>Guide for Determining Mass per Unit Area of Electrodeposited and Related Coatings by Gravimetric and Other Chemical Analysis Procedures</i>

Corrosion Resistance. The environment to which a plated part is to be exposed should be a part of any definition of corrosion resistance. Problems arise in testing a part in its intended environment in part because of the long time period required. In many plating processes, corrosion resistance is directly proportional to the plate thickness, so a specification on plate thickness is a much faster method of indirectly measuring corrosion resistance. In specifying a corrosion resistance requirement for the production of plated goods, accelerated tests are used especially if plate thicknesses cannot be related to corrosion protection.

Corrosion test methods, useful for comparison purposes, usually have no correlation to actual corrosion resistance in atmospheric exposure.

Salt Spray Tests. One of the older accelerated corrosion tests is the salt spray test (40). Several modifications of this imperfect test have been proposed, some of which are even specified for particular applications. The neutral salt spray test persists, however, especially for coatings that are anodic to the substrate and for coatings that are dissolved or attacked by neutral salt fog. For cathodic coatings, such as nickel on steel, the test becomes a porosity test, because nickel is not attacked by neutral salt fog. Production specifications that call for 1000 hours salt spray resistance are not practical for quality acceptance tests. In these cases, the neutral salt spray does not qualify as an accelerated test, and faster results from different test methods should be sought.

The reproducibility of test results between labs using the neutral salt spray tests has not been consistent, but the repeatability, within one lab, is better, and the test has value in comparing variations in coating systems. Correlation of hours of exposure in the salt spray test to actual performance of the plated part in service, even in marine atmospheres, is not consistent and usually avoided. A classic example is that cadmium deposits outlast zinc deposits on steel in salt spray tests and clean marine atmospheres, yet zinc outlasts cadmium when exposed to real, industrial atmospheres, because of the presence of sulfur-bearing corrodents in industrial environments. An important variable in salt spray testing is the position of the surface to be tested. Whereas the surface of test panels is specified to be 15–30° from the vertical (40), when salt spray testing chromated zinc-plated specimens, this range has appeared excessive (41).

The CASS Test. In the copper-accelerated acetic acid salt spray (CASS) test (42), the positioning of the test surface is restricted to 15 ± 2°, and the salt fog corrosivity is increased by increasing temperature and acidity, pH about 3.2, along with the addition of cupric chloride dihydrate. The CASS test is used extensively by the U.S. automobile industry for decorative nickel–chromium deposits, but is not common for other deposits or industries. Exposure cycle requirements are usually 22 hours, rarely more than 44 hours. Another corrosion test, now decreasing in use, for decorative nickel–chromium finishes is the Corrodkote test (43). This test utilizes a specific corrosive paste combined with a warm humidity cabinet test. Test cycles are usually 20 hours.

The Electrolytic Corrosion Test. Also developed for use on nickel–chromium and copper–nickel–chromium decorative automobile parts is the electrolytic corrosion (EC) test (44). Plated specimens or parts are made anodic in a corrosive electrolyte under controlled conditions for 2 min, and then tested for penetration to the substrate.

Adhesion Tests. Adhesion is a measure of the force required to separate the plated coating from the substrate (see ADHESIVES). The bond can be metallurgical, mechanical, or both. When plating metal-on-metal, with good metallurgical bonding, the adhesive strength is greater than the strength of the weaker metal. Plating on surfaces is an atom-to-atom process. Even smooth-appearing surfaces can be relatively rough on a submicroscopic basis, resulting in a physical keying effect that adds to the adhesion. This microkeying is the primary source of mechanical adhesion. Plating on plastics and other nonconductors depends on this microetch for adhesion. Quantitative tests have been developed for coatings on plastics (45), but not to a significant extent for metal-to-metal systems. The quantitative systems in limited use are not commonly used in production plants, but have been used to aid in process development (46–48). For quality control purposes, qualitative tests for metal-to-metal adhesion (49) are usually adequate. The adhesion of some plated metal parts is improved with baking for 1–4 hours at relatively low (120–320°C) temperatures (31). Many of the qualitative adhesion tests vary with the plate thickness. Adhesion appears better for thinner deposits. A specified plating thickness should be a requirement for adhesion testing.

Ductility Tests. The ductility of plated metals differs considerably from the corresponding thermally cast metals. Additionally, ductility which is an important property if parts are to be deformed after plating, varies with the chemical composition of the plating solution, as well as the operating conditions of a given plating process. Ductility can also be important when plated parts are stressed in use. Some metal deposits have coefficients of

thermal expansion significantly different from the substrate. Brittle deposits can crack with modest temperature changes.

Ductility of plated coatings is often measured by bend tests (50), if the plated part can be bent. If not, a foil can be prepared by plating on a nonadherent substrate, removed, and then subjected to a bend test (51). Foils are also tested by conventional tensile testing machines where the foil is pulled apart. Results of these tests have been shown to be affected by the proportionality of specimen thickness to its length and width. Better data are obtained when this is considered. A minitensile testing machine has been described (52) and is commercially available. Significant ductility differences have been found when foils are still attached to substrates. Ductility of foils is also measured by either mechanical or hydraulic bulge-testing techniques. Data obtained by bulge testing vary with plate thickness (53) as do most of the deposit properties. The ASTM B8 Committee has conducted interlaboratory tests on a micrometer bulge-testing instrument.

Hardness Tests. For thin coatings such as electrodeposits, indentation resistance is measured on a microscopic scale, and hardness then becomes microhardness. Hardness (qv), generally related both to wear resistance and to ductility, is often a specified characteristic for plated coatings. Hardness of plated metals is higher than thermally cast metals, and varies with the additives, chemical composition, and operating conditions of each plating solution. Measurements are made by forcing an indenter of specific shape into the test material with a given force, for a given time. For relatively thin-plated coatings, low forces are required, and microhardness values vary according to the force. Because of this, the force used has to be specified along with the microhardness number and the indenter style (54). Proper expression for a microhardness value of 450 using a Knoop indenter and 0.981 N (100 gf) is 450 HK₁₀₀.

The accuracy of microhardness testing has been questioned: a wide range of values appears in the literature for plated deposits, especially in hardness extremes. ASTM B8.10 is involved in interlaboratory testing to define the precision and bias of the Specification B576 Microhardness of Electroplated Coatings (55,56).

Porosity Tests. A pore is defined as a void that extends through the deposit to the substrate. Porosity should also be defined by size. Very fine porosity is present in all deposits; it is only when the porosity is large enough to admit water and corrosdents that porosity becomes significant. Large enough to see, porosity is called pitting. Pits do not always extend through the deposit. Pitting is an intermittent, albeit common, problem. Excess porosity results in decreased service life. Porosity tests for gold, as used for electrical contacts (see ELECTRICAL CONNECTORS), are available (57,58), as are porosity tests of deposits over copper and copper alloys (59), and a general guide to porosity test selection (60).

Hydrogen Embrittlement. Hydrogen is co-deposited with the plate, and atomic hydrogen is absorbed into the surface of the metal during direct cleaning, acid treatments, plating, stripping, and corrosion. If not removed, the atomic hydrogen can have serious embrittling effects on the base metal. This effect is more pronounced with hardened, high strength steels and stainless steels, especially when the hardness is over about 40 HRC, or lower if cold-worked (29,61,62). The mechanism of embrittlement is complicated and not well understood. Relief is accomplished by baking. Generally, the harder steels require longer baking times (63), up to 23 hours at 190°C. Many specifications call for baking to take place within 3–4 hours of plating, although there is some disagreement (64).

Analysis methods for hydrogen absorbed in the deposit have been described (65), and instruments are commercially available to detect hydrogen in metals. Several working tests have been devised that put plated specimens under strain and measure the time to failure. A method for cadmium-plated work has been described (66) as has a mechanical test method for evaluating treatments on AISI 4340 Steel (67). Additional information on testing for hydrogen embrittlement is also available (68).

Fatigue. The phenomenon leading to fracture of a part under cyclical stresses at values less than the tensile strength of the part, is called fatigue. Fatigue fractures are progressive, and grow with the cyclical loads. Plating can have a significantly adverse effect on the fatigue strength (endurance limit) of a part. The greatest loss occurs when the deposit has a high tensile stress. Because a crack cannot propagate in a compressively stressed surface, deposits having higher compressively stresses have much less effect on fatigue life losses. Shot peening the surface before plating also minimizes fatigue loss. Fatigue tests are run on specially treated test specimens in machines that apply cyclical loads. One type of machine uses round test bars that are rotated under adjustable loads. The number of cycles until failure occurs (N) is plotted against the load (S) to obtain an S–N diagram. Unplated control specimens must be run alongside the plated specimens. Fatigue strength loss is an important consideration when plating parts subject to high reciprocating forces. In the aircraft overhaul and rebuilding industry, the effects of plating on fatigue strength are a prime concern. Premature failure of critical parts can be catastrophic (see also FRACTURE MECHANICS; MATERIALS RELIABILITY: NONDESTRUCTIVE TESTINGS).

Internal Stress. Found to some degree in all plated metals, internal stress is either compressive or tensile in nature. It is compressive if the deposit tends to expand, and tensile if the deposit tends to contract to relieve the stress. High tensile stresses can lead to cracking and flaking as well as to serious reduction in life. High compressive stress can lead to blistering when the compressive stress exceeds the adhesion strength. High stress of either nature can lead to warpage where substrates are thin.

Internal stress varies with the deposited metal, the basis material, thickness of deposit, conditions of preparation, operating conditions, and chemical composition of the plating solution. When plating on a dissimilar metal, initial stresses in the first few layers of deposit can be very high perhaps because of epitaxy or the tendency of the deposit to try to match the grain boundaries of the substrate. Other causes for internal stress are not completely explained, but theories often cite the possible effects of minute inclusions of impurities, solid and gaseous, within the deposit or within the grain boundaries, and changes in the parameters of the crystal lattice.

Internal stresses in electrodeposited chromium, nickel, and iron are usually tensile; in zinc, lead, and cadmium these are compressive. Copper deposited from an acid sulfate bath has low tensile stress; copper deposited from a copper cyanide solution has high tensile stress. Additives can be used to change the nature of the stress. Chromium, as deposited from hexavalent chromium systems, develops high tensile stress during plating that can exceed the tensile strength. In such cases, the deposit cracks while plating, and subsequent plating fills in the crack. The cycle repeats as the thickness increases and results in a deposit that can vary from microcracked appearance to macrocracked; where the cracks are completely filled in, it becomes essentially crack-free.

Several methods have been proposed and used to measure internal stresses (69). Most are based on measuring the deformation of a flexible cathode after plating. The cathode is either masked and plated on only one side, or one side is masked after plating on both sides and the plating stripped chemically to produce the same effect. The spiral contractometer (70) utilizes a flat cathode wound in the shape of a helix connected to a dial through a system of gears. The helix is coated with a thin flexible insulating plastic on the inside (71). The rigid strip method uses a flexible metal strip, usually in a nonconductive fixture, to hold the strip in place and to prevent plating on the back side. Methods that involve plating on strain gauges have been described (72), and more recently, a strain gauge method was used to monitor changing stresses. The actual stress in a plated part has been measured by x-ray techniques. However, correlation of internal stress values using different test methods is difficult.

Internal stress values, as measured by deformation, are also influenced by the differences in thermal expansion between the deposited metal and the cathode, especially when plating solutions are used at elevated temperatures. Where the deposit has a higher thermal coefficient of expansion than the substrate, the deposit shrinks more on cooling and adds a tensile stress component to the overall stress. In the reverse case, a compressively stressed component is added. Depending on the relative thermal properties, the same plating solution may be tensilely stressed on one substrate and compressively stressed on another (73). To obtain the stress of the deposit independent from the thermal properties of the substrate, the spiral contractometer can be effectively used by measuring both the initial and final deflection values at the operating temperature of the plating solution. This is termed intrinsic stress of the deposit.

Environmental Aspects

No topic is as important to the plating industry as waste management, which has become the key to economic survival (see WASTES, INDUSTRIAL). Waste management includes recycling (qv), waste reduction (qv), and waste treatment and disposal. Metal finishers (electroplaters) are regulated under a Federal EPA program, National Pollutant Discharger Elimination System (NPDES), often administered by the State. EPA limits on contaminants in plating plant effluents are shown in Table 2. Distinctions are made for established and new plants. State and local authorities have the option of enforcing these limits or establishing more restrictive ones.

Table 2. Pretreatment Standards,^{a, b} mg/L

Parameter	Existing source ^c		New source ^c	
	1 Day	30 Day avg	1 Day	30 Day avg
	<i>Discharges to a POTW^b</i>			
Cd	0.69	0.26	0.11	0.07
Cr	2.77	1.71	2.77	1.71
Cu	3.38	2.07	3.38	2.07
Pb	0.69	0.43	0.69	0.43
Ni	3.98	2.38	3.98	2.38
Ag	0.43	0.24	0.43	0.24
Zn	2.61	1.48	2.61	1.48
CN				
total	1.20	0.65	1.20	0.65
amenable to titration	0.86	0.32	0.86	0.32
TTO ^d	2.13		2.13	
	<i>Discharging to a stream</i>			
TSS ^e	60	31	60	31
oil and grease	52	26	52	26
pH	6–9	6–9	6–9	6–9

^a Final as of July 15, 1986.

^b POTW = publically owned treatment works.

^c Values given are maximum values.

^d TTO = total toxic organics.

^e TSS = total suspended solids.

The Resource Conservation and Recovery Act (RCRA), is the most comprehensive pollution control law for metal finishers. One requirement assigns cradle to grave responsibility for waste to the generator, ie, the plater. Unless the waste is made into another form or recycled as another product, the original generator is still responsible for future treatment or liability costs. Other requirements of RCRA include a generator's license, proper waste containers, labels, and manifest shipping documents, licensed haulers, record maintenance, and an annual waste minimization report. This act was expanded in 1984 and now includes everyone that generates more waste than 100 kg mo. The Clean Water Act was sufficiently strengthened by the Water Quality Act of 1987 (74).

For the foreseeable future, the plating industry has to expect tighter restrictions and more regulation of recovery and recycling operations (see also REGULATORY AGENCIES). Solid wastes, the sludge from waste treatment processes, have to pass stringent leaching tests to be allowed in landfills, and costs for the disposal of solid wastes are increasing dramatically. More recent legislation, the Clean Air Act of 1990, is concerned with air pollutants. This act restricts chromium in air exhausts from chromium-plating and chromic acid anodizing plants to 0.01–0.03 mg m³. The Pollution Prevention Act of 1990 stresses reduction of chemicals that could enter the waste stream before treatment.

Waste Reduction. In 1991, EPA started a voluntary program for industrial plants to reduce releases and off-site transfers of 17 specific chemicals, including

- benzene
- Cd Cd compounds
- carbon tetrachloride
- chloroform
- Cr Cr compounds
- CN-compounds HCN
- Pb Pb compounds
- Hg Hg compounds
- methylene chloride
- methyl ethyl ketone
- methyl isobutyl ketone
- Ni Ni compounds
- tetrachloroethylene
- toluene
- 1,1,1-trichloroethane
- trichloroethylene
- xylene

The target was a 50% reduction by 1995. The goal is pollutant prevention rather than treatment at the end of the pipe. Substitutes are being sought for the heavy metals on the list, and in some applications, substitution can be done without loss in performance. For example zinc-alloy plating is replacing cadmium in some applications. Additionally, other zinc plating solutions have been replacing cyanide zinc baths. In other cases, no suitable substitute may exist.

The main sources of waste from a plating line come from rinse waters, process solution dumps, solution losses from leaks, spills, purifying treatments, and filter changes. Waste reduction occurs with the use of more dilute process solutions, less rinse waters, fewer solution dumps, less filtering, and fewer purification treatments, each of which can be contradictory on many plating lines. For example, because most process solutions act as effective contaminants in the following process solution, poorer rinsing results in shorter process solution life and more frequent purification treatments say nothing about quality. The amount of process chemicals getting into the waste stream can be reduced, however, by longer drain times over the process tank, lower surface tension, more dilute process solutions, misting rinses as parts are being withdrawn, and re-use of the saver-rinse.

Closed Loops and Zero Discharge. A few plating lines have been operated for a period without generating any waste, but not for sustained periods. Zero discharge, a goal being advertised, is not usually attainable. Waste leaves the plant in some form, eg, in sludge, ion-exchange (qv) resins, or filter cakes. Zero discharge often refers to liquid discharges only.

Minimizing Solution Dumps. Process solutions usually have a finite life. This life is extended when extra care is taken to reduce drag-in of contaminants, reduce breakdown products, and avoid reuse of drag-out chemicals unless purified before being returned to the process tank. Another common method of extending the life of electroplating solutions involves purification. The bulk of solution dumps come from dumping cleaners and acids, not electroplating solutions. Methods to increase the tank life of these preplate solutions can have a significant impact on waste prevention in most plants.

Some plants use a bleed-and-feed process where a calculated quantity is withdrawn to waste treat each day and replaced with an equivalent amount of fresh solution. Using cleaners that reject oils can be effective if proper equipment is available. Stripping racks and reject parts off-line helps extend the

life of cleaners and acids. Ultrafiltration (qv) of cleaners, in limited use, is gaining interest. This technique has the capability to remove oils and greases from some cleaning solutions. Improved rinsing between cleaners and acids extends acid life significantly. Inhibitors reduce acid attack and the buildup of dissolved metals in pickle tanks. Continuous filtration (qv), not commonly in use on cleaners and acids, is helpful, packed with activated carbon, filters can keep acid tanks free of oil thus extending the lifetime.

Waste Treatment Methods. Technical and legal specialized knowledge and experience are needed to ensure that goals are to be met when selecting waste treatment methods. A support industry concentrating on waste treatment process and equipment has developed. Table 3 lists several established waste treatment methods commonly used.

Table 3. Common Waste Treatments^a

Method	Treating chemicals	pH	Reaction time, min	Comments
Cr ⁺ reduction, acid	1. Na ₂ S ₂ O ₃ 2. SO ₂ 3. FeSO ₄ ·7H ₂ O	3.0–3.5	30	Fe ²⁺ salts produce four times more sludge than Cr ⁺
CN [−] oxidation CN [−] to CNO [−]	1. Cl ₂ 2. 15% NaOCl 3. Ca(ClO) ₂ ·4H ₂ O ^b	11	60	pH critical to prevent toxic gas
CNO [−] to CO ₂ ⁺ and N ₂ metal precipitation hydroxide	1. Ca(OH) ₂ 2. NaOH	8.5 7.0–10.0	60 30	optimum pH varies with metal to be removed; NaOH produces 1–3 of sludge
sulfide	FeS Na ₂ S	8.0–9.0	20	for polishing after hydroxide precipitation
carbonate	NaHCO ₃	variable	30	good for Pb, Cd, and Ni removal
insoluble starch, xanthate		7.0	fast	polishing after hydroxide; works with some complexing agents present
oily waste demulsification	1. H ₂ SO ₄ 2. polyelectrolytes 3. alum 4. CaCl ₂	3.0 3.10–5.0 3.0 3.0	10–20	removal depends on emulsifiers and oils
ultrafiltration				superior technology

^a Ref. 75.
^b Same as for CN[−] and CNO[−].

Chromium-bearing solutions have to be reduced to trivalent chromium before the metal can be precipitated as the hydrous oxide (hydroxide) by adjusting the pH to the point of minimum solubility. Cyanide also has to be destroyed before metals can be precipitated. Because the pH at which minimum solubility occurs varies with a particular metal, it can be necessary to adjust the pH for each metal, filtering at each change. The initial pH range is usually 9.0–9.5 and varied from there, as required. Metal salt solutions containing complexing agents present special problems, because these salts may stay in solution at all pH ranges. In some cases, calcium salts such as lime can displace the metal from the complex at high pH; in other cases, electrolysis may help. For some strongly complexed systems, such as those found in some metal stripping chemicals, the metal bond cannot be removed, and incineration may be required (see INCINERATORS). There is a strong trend away from the use of strong complexing agents in the plating industry and elsewhere.

Once a metal has been precipitated, the solution is pumped to a clarifier or some type of settling tank where the metal-bearing sludge is separated, usually ending up in a filter cake. Drying (qv) or dewatering (qv) of the cake results in less weight and lower shipping costs.

There are other methods for waste treatment. Cyanide may be treated by ozone (qv) oxidation, peroxide formaldehyde treatment, and electrolytic oxidation. A procedure using sulfur dioxide air oxidation has also been proposed to destroy cyanide and permit removal of most of the metal–cyanide complexes in copper-plating solutions (76). Sodium borohydride has been used to reduce metal ions directly to metals. Combination methods involving some form of concentrating treatment, such as electrodialysis (see ELECTROSEPARATIONS) reverse osmosis (qv) or regenerants from ion-exchange (qv) resins, along with electrowinning, can produce a more saleable metal waste. Electrolytic processes are used with specialized, high surface area cathodes in proprietary equipment where the cathodic material can be salvaged directly, or through the supplier of the equipment. Ion-exchange units are handled similarly; the waste or regenerant can be reused in-house or exchanged with the supplier in some cases.

Recycling vs Waste Treat. It is becoming increasingly difficult to dispose of the hydroxide sludges generated from the treatments in Table 3. Thus there is a trend toward recycling or recovery. Sources that can use metal-containing sludges as feedstock are becoming more common (77). Except for precious metal sludges, the value of metal-bearing waste treatment sludge is not significant, and the plating plant incurs the costs of shipping and treatment. In the interest of avoiding future liability, recycling is expected to grow. The recovery or recycling of metal-bearing sludges or solutions and the value of scrap metal from electro-winning treatments is enhanced by higher concentrations and purity. Mixed metals are less desirable than single metals; thus, it may be advantageous to treat metal waste streams separately.

Plating Bath Formulations

Formulations of plating baths can be flexible in some systems and very sensitive to variation in others. Many of the more recent changes have resulted from waste treatment and safety requirements. Besides the ability to deposit a coating having acceptable appearance and physical properties, the desired properties of a plating bath would include: high metal solubility, good electrical conductivity, good current efficiencies for anode and cathode, noncorrosivity to substrates, nonfuming, stable, low hazard, containing, low anode dissolution during down-time, good throwing power, good covering power, wide current density plating range, ease of waste treatment, and economical to use. Few formulas have all these attributes. Only a few plating solutions are commercially used without special additives, but chemical costs often constitute a relatively low percentage of the total cost of electroplating. Additives are used to brighten, reduce pitting, or otherwise modify the character of the deposit or performance of the solution. Preferred formulations are normally specified by the suppliers of the proprietary additives.

Plating Bath Purification. Purification, often needed once a plating bath is made, is used periodically to maintain the plating solutions. Alkaline zinc plating solutions are sensitive to a few mg/L of heavy-metal contamination, which can be precipitated using sodium sulfide and filtered out.

Nickel plating solutions may contain excess iron and unknown organic contaminants. Iron is removed by peroxide oxidation, precipitation at a pH of about 5, then filtered out. The more complex, less water-soluble organic contaminants along with some trace metals are removed with activated carbon treatments in separate treatment tanks. About 5 g L of plating-grade activated carbon is mixed in the plating solution for at least 1–2 hours, usually at warmer temperatures.

Another common purification treatment used both on new and used plating solution is dummying. Heavy-metal impurities are removed by electrolyzing, usually at low current densities, using large disposable steel cathodes. Good agitation and lower (acid) pH speed the process. In alkaline plating solutions, dummying is often more effective at higher pH; eg, zinc is dummied out of cyanide copper plating baths faster after caustic additions.

Testing and Control. Analysis and testing are required whenever a new plating solution is made up, and thereafter at periodic intervals. The analyses are relatively simple and require little equipment (78–80). Trace metal contaminants can be analyzed using spot tests, colorimetrically, and with atomic absorption spectrophotometry (see TRACE AND RESIDUE ANALYSIS). Additives, chemical balance, impurity effects, and many other variables are tested with small plating cells, such as the Hull cell developed in 1937 (81,82).

Frequency of analysis and testing should be adjusted so that corrections can be held to 10⁰ º or less from established concentrations. Sophisticated automatic analytical machines are commercially available.

Individual Plating Baths

Cadmium. Compositions of cadmium plating baths are shown in Table 4. Whereas cadmium provides better corrosion resistance to steel and other substrates than zinc when exposed to marine environments, zinc is better in industrial, sulfur-bearing atmospheres. Resistance of cadmium is improved with chromate conversion coatings; bright, yellow iridescent, olive drab, and black finishes are used. Cadmium is readily solderable, provides lubricity on threaded fasteners, and has been used as a high temperature protective coating in the aircraft industry when diffused into nickel plating. Olive-drab chromated cadmium plate is used on nickel-plated aluminum electrical connectors to extend salt spray resistance. For diffusion processes, for low hydrogen embrittlement, and for plating directly on cast iron, cadmium–cyanide plating solutions are used without brighteners; in other applications, brighteners produce attractive lustrous deposits.

Table 4. Cadmium-Plating Solutions^a

Parameter	Cyanide	Acid sulfate	Fluoborate
Cd metal, g L	20–30	15–30	75–150
NaCN, g L	90–150		
Na ₂ CO ₃ , g L	30–60		
H ₂ SO ₄ , g L		45–90	
NH ₄ BF ₄ , g L ^b			60–120
NaOH, g L	10–20		
pH	12.5–13.5	very acid	2.5–3.0
anodes	Cd and steel	Cd and graphite	Cd
temperature, °C	15–40	18–23 max	10–40
current density, A m ²	50–900	300–800 ^c	100–600

^a Addition agents for all three plating solutions are as required.
^b Some baths contain 20–30 g L NaBF₄ and 20–30 g L H₃BO₃ instead of NH₄BF₄.
^c Current densities from 100–200 A m² are used for barrel plating.

The future of cadmium plating is in question because of cadmium toxicity (83). Cyanide-free cadmium plating systems have experienced some growth. Acid cadmium, based on cadmium sulfate compositions, is replacing some cyanide baths in the United States. The fluoborate cadmium is reported in use in the UK, especially in barrel plating (84). Cadmium plating is covered by ASTM (85), U.S. government, and ISO specifications (86).

Chromium. Applications of chromium plating can be separated into two areas: hard chromium, also called functional, industrial, or engineering chromium, and decorative chromium. The plating bath compositions may be the same for both. In most cases, the differentiating factor is plate thickness. Decorative chromium is usually less than about 1 µm; hard chromium can be from about 1 µm to 500 µm or more.

Chromium is conventionally deposited from chromic acid solutions containing at least one anionic catalyst, which is usually the sulfate ion. The weight ratio of chromic acid to catalyst is important and, for sulfate-catalyzed solutions, is maintained about 100:1. Formulations and conditions for operating hard chromium plating solutions are shown in Table 5.

Table 5. Chromium-Plating Solutions^a

Parameter	Conventional baths		Cocatalyzed baths	
	Range	Typical	Range	Typical
CrO ₃ , g L	150–400	240–260	150–400	150–180
SO ² ₄ , g L	1.25–5	2.4–2.6	0.6–1.3	0.9–1.0
SiF ² ₂ , g L			0.3–0.8	0.5–0.6
CrO ₃ :SO ² ₄ ratio	80–120:1	90–110:1	150–200:1	170–180:1
anodes	Pb-7 ^o Sn or Pb-6 ^o Sb		Pb-7 ^o Sn	
current density, A m ²	400–4000	1000–3000	400–4500	1000–3500

^a All decorative chromium baths are run at 38–43°C; hard chromium baths at 40–60°C.

Chromic acid-based plating solutions differ from the other common metal-plating baths in that chromium solutions have poor current efficiencies, poor covering power, and poor throwing power, as well as a need to use much higher current densities to plate. Modifications developed in attempts to improve these weaknesses incorporate a second catalyst, and are called cocatalyzed baths. These have improved cathode efficiency: cocatalyzed systems can produce cathode efficiencies of 16–17^o vs the 8–12^o of the conventional baths. The second catalyst is often fluoride, or more commonly, the fluosilicate ion. Proprietary baths are also marketed having self-regulating, chromic acid-to-catalyst ratios for ease of operation. A problem exists for the fluoride-containing catalysts in that the plating solution severely etches any steel areas exposed to the bath that are not being plated such as deeply recessed areas, inside surfaces of hollow cylinders, and other low current density areas. These nonplating areas have to be stopped-off with protective materials such as waxes, tapes, or other masking compounds. Newer proprietary compositions are available that claim nonetching, fluoride-free, co-catalyzed systems having efficiencies up to about 25^o.

Compositions, as shown in Table 5, can be quite flexible as long as the catalyst ratio is well controlled. High chromic acid baths produce less burning and have better conductivity but have lower current efficiency and a higher drag-out loss. Lower chromic acid baths exhibit better coverage in low current areas. Low catalyst ratio (high sulfate) baths have decreased coverage, but smoother deposits. High sulfate is reduced by barium carbonate precipitation. About two parts of barium carbonate is needed to precipitate one part of sulfate in a slow reaction that takes several hours to complete. Low sulfate (high ratio) also gives better coverage, and plating rate is decreased. Temperature control is important in all chromium plating baths, and agitation is required in most tanks to avoid temperature stratification. Higher temperatures improve conductivity of the solution, decrease burning, and decrease covering power. Lower temperatures produce slower plating rates.

Anodes used in chromic acid plating baths are insoluble lead–tin or lead–antimony alloys. Chromium anodes are not used because they dissolve too quickly. The lead–alloy anode has a secondary benefit that, under correct conditions, trivalent chromium is reoxidized at the anode surface, and a damaging buildup is avoided. Lead–alloy anodes develop a dark brown oxide coating when working properly. If left idle or without sufficient current flow, the anode forms a chromate coating that decreases the current flow and is not capable of chromium-reoxidation. When this yellow coating is observed, it may be necessary to remove and clean the anode. Caustic–gluconate solutions are often used for anode cleaning. Anode current density is important to maintaining the proper oxide; too high or too low are both detrimental. Because throwing power is so poor, anode placement and anode-to-cathode distances are very critical in obtaining sufficient plate thickness distribution. Anodes that conform to the shape of the part give the best results. Anodes should be shorter than the work. Anode-to-cathode spacing for conforming anode arrangements is usually about 10 cm; for internal anodes in hollow cylinders smaller spacing, 1-2 cm allows for more anode surface.

Impurities. Chromic acid plating solutions are affected by metallic impurities that usually are accompanied by chromium reduction increasing electrical resistance. Iron is a common impurity that increases rapidly when the chromium plating tank is used as the reverse current chromium etch tank. Eventually, chromium plating solutions have to be replaced because of high iron content. Reverse etching is used to obtain good adhesion; over-etching should be avoided. A separate chromic acid etch tank is used to extend the life of the plating solution. Trivalent chromium can be treated by electrolytic oxidation treatments at temperatures of about 80°C and high (5000–5500 A m²) cathode and low (2000 A m²) anode current density. Limits of trivalent chromium that can be tolerated vary from 10–60 g L. Treatment using the electrolytic porous pot is also used to purify chromium solutions (87).

Other purification treatments include: cation-exchange treatments on diluted solutions followed by evaporation back to working bath strength; and electro dialysis directly in the working solution (88). Fluoride ion excess can be treated using aluminum foil but high (10 g L) aluminum can completely inhibit chromium plating in cocatalyzed baths. High chlorides, which reduce plating range and increase anode deterioration, can be removed by precipitation with silver oxide salts or by controlled dummying using an anode current density of about 210 A m².

Good ventilation is required for chromium plating. Chromic acid plating solutions having low current efficiencies produce large quantities of chromic acid-laden spray requiring good air scrubbers (see AIR POLLUTION CONTROL METHODS). Certain wetting agents form foam blankets on chromium-plating solutions and reduce spray losses, but some wetting agents contribute to pitting in hard chromium with plate thicknesses over about 25 μm.

Decorative Chromium. The thin (0.25 μm) bright chromium plate deposited over bright nickel gives a blue-white color, and was long thought of as only a tarnish preventative. This outermost coating or first barrier, plays a significant role in retarding corrosion, however, even though it has a tendency to crack. Chromium-plated goods replated with a second layer of chromium, when the first layer was found thinner than specified, were found to have better corrosion resistance. In the second layer, many microcracks (MC) were found that extended down to the nickel plate. Moreover, less corrosion appears on chromium-plated test pieces at the ocean beach than on parts placed away from the shore, and plated ware on the front of automobiles are generally less corroded than parts on the rear. In both of these cases, the plated surface shows a microporous (MP) effect produced by the sand blasting. Additionally, increased protection has been observed when chromium is plated over satin nickel where the satin-effect is caused by codepositing fine, nonconducting particles with the nickel. In this case, it was shown that the chromium was microporous and did not plate over the tiny particles. A minimum number of discontinuities is required for MC chromium, (300 cracks cm) and MP chromium (10,000 pores cm²) to be corrosion resistant. Automotive and ASTM specifications (89) require either MC or MP chromium on extended severe service parts.

Trivalent Chromium. Use of trivalent chromium plating for decorative applications is continuing to grow. These materials have performed well in corrosion tests (90). There are two types of proprietary trivalent chromium plating baths being marketed. These differ in anode systems and are compared in Table 6. Suppliers of trivalent chromium-plating processes claim several advantages over chromic acid baths (91). Thick deposits up to about 25 μm appear to be attainable, but the use of trivalent chromium for hard chromium applications is not commercially practiced to any known extent as of this writing.

Table 6. Trivalent Chromium-Plating Systems^{a, b}

Parameter	Single cell	Double cell
anode material	carbon	Pb–7% Sn in diaphragm cell
maximum thickness, μm		
room temp	1.3	0.25
high temp	25 or more	
plating rate, μm min		
room temp	0.13	0.10 ^c
high temp	0.18	
deposit ^d	MC and MP	MP
passivity of nonplated areas	needs post dip	needs post dip

^a Ref. 91.

^b Both cell types utilize 5–25 g L Cr at a pH 2.3–40 and a temperature from 20–50°C under mild air agitation. Current densities run from 430–1600 A m², the rectifier voltage is from 4–15 V, and the anode to cathode ratio is 2:1.

^c Average value.

^d MC microcracked; MP microporous.

Specialty Chromium-Plating Baths. Chromic acid baths using sodium chromate and sodium hydroxide to form a tetrachromate (92) have had limited use. Porous chromium is used in lubricated wear applications, and is made by chemically etching regular chromium plate, sometimes with light grinding after the etch. Black chromium is used on solar collector surfaces (see PHOTOVOLTAIC CELLS; SOLAR ENERGY). Baths are sulfate-free, and include fluosillicic acid or acetic acid (91).

Standard practices for chromium plating (93) and specifications for hard chromium (94) and decorative chromium (89) have been published.

Cobalt. Cobalt plating is little used because most applications can be satisfied using nickel at considerably less cost. Cobalt has been used to coat tungsten carbide components that needed to be brazed and exposed to high temperatures in use; and as a barrier coating to inhibit migration of copper into gold. Cobalt is sputtered onto computer memory disks for its magnetic properties (see INFORMATION STORAGE MATERIALS). Cobalt is used in nickel electroforming solutions to increase strength and hardness. Cobalt plating solutions are so similar to nickel plating solutions in composition (95,96) that in most cases the equivalent cobalt salts can be substituted for nickel salts in published nickel bath formulations. Cobalt anodes dissolve well in sulfamate solutions, and have been used as auxiliary anodes in nickel sulfamate electroforming solutions.

Copper. Copper plating, used as a final finish in some applications, is primarily employed as an undercoat for other deposits. Of the several types of copper baths, the two most popular in the United States are the acid sulfate and the cyanide baths. Acid copper cannot be used directly over

substrates that are attacked by high acidity, or over those on which the copper forms immersion deposits. Immersion deposits usually have poor adhesion to the substrate. For such substrates, an undercoat deposited from a cyanide copper solution is used.

Cyanide copper baths are a target for replacement, and alkaline, noncyanide baths are being advertised as replacements for cyanide copper. Care is needed in choosing a noncyanide copper system to avoid trading one waste treatment problem for another. As of this writing, cyanide copper solutions are still being used for plating the initial deposits on steel, zinc-based die castings, aluminum alloys, magnesium alloys, and other substrates.

Cyanide Copper. Formulations for copper cyanide baths are shown in Table 7. The more dilute, less efficient copper strike bath is a necessary and important step in the overall performance and appearance of the final plated part. Cyanide solutions are known for cleaning ability, which is enhanced using the low efficient, high gassing cyanide copper strike baths. Strike tanks, which have several advantages, are small compared to the following plating tanks, allowing for rapid turnover of the solution through filters to remove particles and soil before the parts reach the large, more expensive plating tank. Strike compositions are modified usually through pH changes for striking various substrates. For use on steel, additional hydroxide improves conductivity and helps protect steel anode containers and tanks from corrosion. High potassium hydroxide is said to reduce anode efficiency, which can be beneficial in strike baths. Copper strike solutions tend to increase in copper and eventually need dilution. If steel anodes are used along with copper anodes, rate of copper buildup can be reduced, but the rate of carbonate formation is increased. Dilute acids have been used to lower the pH of cyanide solutions, but the safety of this practice is questionable. Contact of the work with the d-c power supply is often required before entering the strike solution.

Table 7. Cyanide Copper-Plating Solutions^a

Parameter	Solution		
	Strike	Rochelle	High speed
Cu metal, g L	15–22	22–36	56–71
CuCN, g L	21–31	31–51	79–100
KCN			
total ^b	31–70	55–89	130–165
free ^c	10–15	12–18	20–25
KOH, g L	3–18 ^d	15–25	15–25
Rochelle salts, g L	10–20 ^e	30–45	35–55
Na ₂ CO ₃ , g L ^f	10–15 ^e	15–30 ^e	15–30 ^e
temperature, °C	42–49	54–60	60–70
current density, A m ²			
cathode ^g	160–320	up to 270	up to 380
anode ^h	50–100	110–160	110–160

^a Wetting agents and other additives are added as required.

^b Total KCN = 1.4 (CuCN) + free KCN; total NaCN = 1.1 (CuCN) + free NaCN.

^c Free KCN = surplus KCN over that required to complex the CuCN; the titratable KCN. NaCN can be substituted on an equimolar basis with the CN content. Current density range is less and throwing power is less but NaCN costs less.

^d KOH in strikes for zinc die casting is kept below about 3 g L; for Cu strike on zincated Al, sodium bicarbonate is added to lower pH to about 9.8–10.2.

^e Optional additive.

^f CO² forms in quantity and increases with use. Sometimes CO² is added to new baths to break them in.

^g Current densities cited are average based on racked work, not Hull cell derived.

^h Higher currents polarize anodes; current density is critical.

Rochelle salts produce some grain refinement, reduce the effects of some metallic contaminants, and increase the anode current density range before anode polarization occurs. Plating rates are slower for low metal contents. Rochelle salt copper baths are useful with higher metal contents for buffing or for heat treat stop-off applications. Higher currents produce harder-to-buff deposits; some proprietary additives can extend the current density range. Higher current density and higher carbonate content increase the deposit porosity; additives can be helpful and carbonates should be kept below about 120–150 g L. With increased porosity, thicker deposits should be used for heat treat stop-off. Carbonates, always present in cyanide copper solutions, are beneficial in smaller amounts but detrimental if too high. Carbonate is formed normally from cyanide oxidation, but builds up much more rapidly when using insoluble or inefficient anodes. Carbon dioxide is also readily absorbed from the air to form carbonate salts. Thus air agitation of cyanide copper solutions, a common practice, is not recommended. Carbonate from a sodium-containing copper bath can be precipitated by cooling the solution. In northern climates, solutions are treated by pumping to an outside tank in the winter to freeze out the sodium carbonate. Potassium carbonate salts are too soluble to be removed in this manner, and are precipitated by addition of calcium salts. Acceptable carbonate limits vary widely; better results are obtained below 120–150 g L as sodium carbonate. High carbonate lowers the anode efficiency, which accelerates carbonate formation as well as producing rough deposits. Proprietary products are available to yield improvements in anode corrosion.

High speed formulations allow use of higher currents. These are sometimes also referred to as high efficiency baths. Cyanide-plating baths typically decrease in cathode efficiency with increased current, which accounts for the good plate distribution. High speed baths having higher metal content permit the use of higher currents, and additives extend this range, producing additional grain refinement and brightness. Certain selenium compounds, along with very small amounts of lead or thallium combinations, produce bright deposits. Sulfur compounds and selenium compounds codeposit, and the copper deposit is more active and tends to tarnish (oxidize) readily. Systems that avoid these have shown better resistance to accelerated corrosion tests when overplated with nickel and chromium. Additives are necessary to avoid sulfur codeposition, because sulfur is a common impurity in cyanide salts and plating solutions. Cyanide copper baths are sensitive to small amounts of certain organic contaminants such as oils and greases; small amounts of nonionic and anionic detergents are used to eliminate the smeared or pitted appearance of the deposit. Hexavalent chromium in as small as 1 mg L amounts can cause blisters and dullness. In the presence of trivalent chromium, copper baths have been operated as high as 3–4 g L of chromium. Sulfur-bearing reducing agents have been used, but proprietary sulfur-free products are preferred.

A common problem with cyanide copper is rough deposits much of which comes from anode films that slough off and settle on the work before the film particles can redissolve. Higher free cyanide produces less anode filming, but lowers the cathode efficiency and requires lowering the current. Along with the chemical composition and control, anode current density control, filtration, and agitation are key factors in minimizing roughness. Anode bags of suitable weave and material have been used successfully.

Leveling of copper cyanide plating baths can be obtained through d-c current manipulation (15). Periodic reversal (PR) cycles combined with proper bath composition can change the leveling from a negative value to about 50% based on profilometer readings on a 15-µm deposit. Metallic brighteners and sulfur-bearing additives are avoided in the PR process.

Barrel plating of parts in copper cyanide solutions utilizes various formulations, some weaker, some stronger than the high speed baths. When plating parts that tend to stick together or nest during the barrel rotation, the free cyanide may need to be increased. This may require 35–40 g L free potassium cyanide or more with an equal copper content.

Cathode efficiency and throwing power can vary with chemical composition, additives, operating conditions, and impurities in the cyanide copper bath (97). In production tanks, the cathode efficiency of the so-called high efficiency bath formulations approach 100% only at low (often <50 A m²) current densities. Using more practical currents of 300–350 A m², the cathode efficiency may drop 20%, especially with low agitation rates. It is this drop in efficiency with increased current that gives cyanide copper baths such good throwing power.

Acid Copper. Bath compositions are shown in Table 8. The acid sulfate bath is by far the most widely used copper plating bath, both for plating and for electroforming and electrowinning. The fluoborate baths have been little used in spite of the high current densities possible. Additional information can be found in the literature (98,99).

Table 8. Acid Copper-Plating Baths^a

Parameter	Sulfate			Fluoborate	
	Typical	Average range	High throw	Low	High
Cu metal, g L	57	38–64	15–23	60	120
CuSO ₄ ·5H ₂ O, g L	225	150–250	60–90		
H ₂ SO ₄ , g L	60	30–75	170–220		
Cl [–] , g L	0.05	0.20–0.12	0.05–0.10		
Cu(BF ₄) ₂ , g L				225	450
HBF ₄ , g L				15	30
H ₃ BO ₃ , g L				15	30
pH	<0	<0	<0	1.2–1.7	0.2–0.6
temperature, °C	25	20–45	20–30	18	65
cathode current density, A m ²	300–700	300–700	300–700	800–1400	1400–3800
anodes ^{b,c}		0.02–0.08% PCu		OFHC Cu	OFHC Cu

^a For all baths the anode to cathode ratio is 2:1; the baths are mechanically, hydraulically, or air agitated, and there is continuous filtration.

^b Dynel or polypropylene anode bags are used for both types of anodes.

^c OFHC = oxide-free high conductivity.

A large amount of copper foil is electroformed on a continuous basis from copper sulfate baths for use in the printed circuit industry (see ELECTRONIC COATINGS; INTEGRATED CIRCUITS). Thick smooth, low stressed deposits that are easily buffed are obtained from these relatively simple baths, and applications are found in many engineering and decorative markets. Very bright, highly leveled deposits are plated using proprietary additives. The two main deficiencies are poor throwing power and the inability to plate directly on steel and certain other substrates. Acid copper sulfate is also used to plate thick (175–200 μm) deposits over large nickel-plated rolls, then engraved, in one method, to electroform textile printing screens (see TEXTILES). Plates and rolls have been acid copper sulfate plated for graphic arts and rotogravure use. Bright acid copper sulfate baths are used extensively in decorative plating of plastic trim found on automobiles, appliances, and household wares. Copper can be plated directly over the first metallic coating of electroless nickel or electroless copper. Very low current is used in the acid copper, initially, to avoid burning off the thin electroless deposit. More often, a nickel strike based on a Watt's bath is used to build up the electroless coating, before the acid copper. The good ductility and low internal stress are key features for plating on plastics (100).

Rolls and other relatively simple shapes make use of inert shields and thieves to avoid edge buildup and produce a more even plate thickness. For more complicated shapes having deeper recesses thicker deposits from cyanide copper baths have been used as an undercoat to the copper sulfate deposit. Acid copper baths operate near 100% efficient over a wide current density range. The cathode efficiency is usually slightly less than the anode efficiency, bringing about a slow increase in copper unless drag-out losses are high.

Not all grades of copper sulfate are suitable for electroplating because some contain significant amounts of silica and other materials used to improve the crystallization process and the free-flowing properties of the salt. Sulfuric acid concentration has more effect on conductivity than does the copper sulfate. Low sulfuric acid content produces more high current density (HCD) burn and poorer leveling, low current density (LCD) dullness and gives more nodular deposits. High sulfuric acid has less effect on the deposit but increases anode dissolution. Chloride-ion is important, especially for baths that contain additives and brighteners because chloride inhibits rough and nodular plate, and is necessary for bright deposits. Low chloride can cause dark deposits on edges and other HCD areas, loss of leveling, loss of brightness, pitting, and poor anode corrosion. High chloride causes striations, increased brightener usage, reduced leveling and brightness in bright baths, and nodular, treed, and harder deposits. High chloride can be reduced with zinc dust treatments or precipitation with silver nitrate followed by filtering to recover the silver. Zinc anode balls in a basket have been used with successive immersion and cleaning steps.

High throw formulations, such as those in Table 8, are used on PTH (plated-through-hole) printed circuit boards and in barrel plating where the high acid low copper bath exhibits better throwing power. Acid copper sulfate solutions are tolerant to many metallic impurities. Iron is common, and has little effect in low concentrations; above about 2 g L, iron decreases conductivity and reduces throwing power. Organic contaminants from brightener breakdown; oils, tank linings, and anode bags are removed with activated carbon treatment.

Copper Pyrophosphate. Baths using copper pyrophosphate [10102-90-6], Cu₂P₂O₇, shown in Table 9, are not in wide use in the United States as of this writing. Baths are used for printed circuits, plating on plastic, and electroforming (38). Although a pyrophosphate strike can be used to plate on steel, zinc-based die castings still require copper cyanide strikes for good adhesion. Waste treatment of pyrophosphate copper solutions presents some problems. There is difficulty in precipitating the copper without hydrolyzing all the pyrophosphate. Further, many areas now restrict the amount of phosphate in the waste stream.

Table 9. Copper Pyrophosphate-Plating Baths^a

Parameter	Strike	Typical	Printed circuit bath
Cu metal, g L	9–11	19–30	21–26
Cu ₂ P ₂ O ₇ ·3H ₂ O, g L	25–31	53–84	58–73
K ₄ P ₂ O ₇ ·3H ₂ O, g L	112–205	235–405	270–370
NH ₄ OH, mL L	0.5–1.0	3.75–11.0	2.7–7.5
KNO ₃ , g L	1.5–3.0	3.0–6.0	8.0–16.0
P ₂ O ₇ :Cu ratio	7–8:1	7.0–7.5:1	7.2–7.8:1
pH	8.0–8.5	8.0–8.7	8.0–8.4
temperature, °C	22–30	43–60	49–54
cathode current density, A m ²	60–150	100–800	250–600

^a For all baths the anode to cathode ratio is 2:1, the baths are mechanically or air agitated, and there is continuous filtration.

Copper–Zinc Alloys. Brass, the first alloy to be electroplated, is a popular decorative finish. Plated as an undercoat for nickel–chrome, brass has given good corrosion protection compared to copper, especially on aluminum and zinc substrates. There has been considerable use of brass plating on steel wire and other shapes because of improved rubber bonding characteristics. Several copper–zinc alloys are commercially plated; the more common are yellow brass (ca 30° o zinc), red brasses (10–15° o zinc), and white brass (ca 72° o zinc). Typical bath compositions are shown in Table 10. Decorative yellow brass plating is usually a very thin, flash coating over a bright nickel deposit. Brass plating times may be as short as 30–60 s at 100–200 A m². The brass is passivated in a dilute chromate and lacquered to protect the finish. If brass is to be used for rubber bonding, the thin (0.5–2.0 μm) coating is not chromated or lacquered. Bath compositions for yellow brass, can vary. Control of the yellow color is simplified by the use of ammonia to redissolve the zinc. The cyanide:zinc ratio plays a key role in the proportion of copper–zinc deposited. Other, more alkaline baths have been used with Rochelle salts or other complexing agents to keep all components in solution. Noncyanide brass plating baths have been publicized (101), but there is no significant use as of this writing. Brass plating is well covered in the literature (102,103). High copper alloys resemble bronze.

Table 10. Brass-Plating Solutions

Parameter	Yellow brass		White brass
	Regular ^a	High speed	
CuCN, g L	32	75	10
Zn(CN) ₂ , g L	10	5	60
NaCN, g L ^b	50	125	100
Na ₂ CO ₃ , g L	7.5		40 ^c
NaHCO ₃ , g L	10		
NaOH, g L		45	38
NH ₄ OH, mL L ^d	2.5–5.0		
pH	9.7–10.2		12.5–13.5
temperature, °C	25–35	70	15–20
cathode current density, A m ²	up to 250	100–800	200–800
anodes	70:30 brass	70:30 brass	Zn or white brass, 35° o Cu
Cu in deposit, ° o	70 ^e	70	28

^a 70:30 yellow alloy is controlled by CN:Zn ratio; Zn should be 1 3–1 5 of titratable CN . Bath can be diluted to 65° o for flash plating.

^b Titratable CN is less than total and is used to control bath. Barrel plating solutions are operated at higher CN and Zn levels.

^c Value given is maximum.

^d High ammonia favors Zn; low ammonia gives redder deposit.

^e Red brass, 85:15, is obtained using higher CN :Zn ratio; architectural bronze, 90:10, with an even higher ratio.

Copper–Tin Alloys. Bronze electrodeposits have been used for decorative applications as a final finish, with and without subsequent coloring or antiquing treatments. A clear lacquer helps preserve the deposit. On a limited basis bronze has also been used as an undercoat for nickel–chromium finishers on steel, zinc, and aluminum. In functional applications, plated bronze is used as a stop-off material in the selective nitriding of steel parts. It also finds good use as a nongalling bearing material (see BEARING MATERIALS). One process simultaneously hones and plates the inside diameters of bearing surfaces in a high speed operation. Speculum deposits, when polished to a high luster, resemble silver plate, have good hardness, and resist staining by most foods. These deposits have found use on tableware and household fixtures; speculum plating is not recommended for outdoor use. Typical bath formulations for bronze plating baths are shown in Table 11. Bronze plating is done primarily in cyanide–stannate baths; noncyanide baths appear in the literature (104), but are not in common use. The more common bronze contains about 10–15° o tin; speculum bronze is 40–45° o tin.

Table 11. Typical Bronze-Plating Solutions

Parameter	(1)	(2)	Speculum
CuCN, g L	35	29	11
Cu metal, g L	25	20	8
Na ₂ SnO ₃ ·3H ₂ O, g L	38	35	90
Sn metal, g L	17	14	40
NaCN, g L			
total	54	64	27
free	15	22	16
NaOH, free g L	7.5	10 ^a	16
Rochelle salts, g L		45	
pH	12.6		
temperature, °C	65	65	65
cathode current density, A dm ²	3	2–10	3
Sn in deposit ^b , ° o	10 ^c	12	40 ^d
anodes	Cu ^e	Cu	Cu or dual Cu and Sn

^a Potassium salts, which permit higher currents and better efficiency, are used.

^b To increase Sn in deposit, raise NaCN and decrease NaOH. New bath may need 0.2 g L ammonia to obtain high Sn.

^c Ductile.

^d Brittle.

^e If using bronze anodes, Sn should be less than 10° o and anode current density <1 A dm².

Gold. Electroplated gold is used both for decorative and industrial or engineering purposes. The engineering use has grown rapidly since the 1950s, because of expanding usage in the electronics and computer industries. The important properties of industrial gold are corrosion resistance, low electrical contact resistance, and good solderability. The introduction of small amounts of other metals reduces one or more of these properties, often significantly, eg, as little as 1° o iron increases the electrical resistance of gold over 1000° o. Electronic applications include semiconductors (qv), printed etched circuits, and contacts connectors. Engineering gold is classified into three types based on purity and hardness (105). Type 1, code A is used

for semiconductor components, thermocompression bonding, and where ever high temperature resistance is needed. Type 2 is for general purposes, high reliability electrical contacts, solderability, and wire wrap connections. Type 3 is used for wear resistance, but does not stand high temperatures. Types 1 and 2 are preferred for solderability.

Gold is usually plated over electroplated (or electroless) nickel-plated substrates. The smoothest substrate allows the thinnest gold deposits without porosity. For many substrates, the minimum gold thickness for pore-free deposits is about 2.5 μm. Gold plate thicknesses vary from a flash or wash of 0.5 μm up to about 10 μm depending on application. In electronic uses the range is more often 0.5–1.25 μm. Thicker gold deposits produce brittle solder joints. Decorative gold plate has to be a minimum of about 0.18 μm to be identified as electroplated gold on jewelry. If thickness is 2.54 μm, it can be classified as heavy gold electroplate. A small amount of gold is used in electroforming with substantial thicknesses. Aerospace applications for both thin and thick deposits are newer uses for gold.

There are literally hundreds of gold-plating bath formulations; most of which are proprietary. Gold flash, gold strikes, soft golds, hard golds, bright golds, and golds of all purities and colors account for the wide selection (106). Formulations can be classified as alkaline, neutral, or acidic; typical baths are shown in Table 12.

Table 12. Gold-Plating Bath Compositions

Parameter	Alkaline ^a	Neutral ^a	Acid ^a	Strike
Au metal, g L	2–12	4–16	2–16	0.5–2.0
KAu(CN) ₂ , g L	3–18	6–24	3–24	0.75–3
KCN, g L	15–48			15–90
K ₂ CO ₃ , g L	0–45			
K ₂ HPO ₄ , g L	0–45	0–90	0–100	15–45
KOH, g L	1–30			
pH	11–13	6–8	3–6	8–13
temperature, °C	50–70	25–70	40–70	40–60
cathode current density ^b , A m ²	10–100	20–100	10–100	
anodes ^c	Pt, SS	Pt, SS	Pt	SS

^a Proprietary brighteners and chelators are used.
^b Higher currents require higher gold contents.
^c SS = stainless steel; Pt may be platinized titanium or platinized niobium.

Historically gold baths, even acid gold, were primarily based on cyanide. Gold solutions and rinses are recovered for gold value. Of the noncyanide gold systems, gold sulfite baths which may contain phosphates, citrates, and chelating agents, have been growing in use. Claims are made for better ductility and throwing power, good alloy deposition, and more tolerance to impurities. Baths are proprietary.

Indium. Indium diffuses readily into copper, silver, and lead, increasing hardness and imparting good antifriction properties. Indium plating of copper wire, used in joining aluminum wire to copper wire, has declined with decreased use of aluminum wire in housing. Indium plating is used for bearing materials, and some is used in electronics for contact pads for flip chip bonding. Although indium can be plated from cyanide, fluoborate, and sulfate baths, the standard bath recommended is based on sulfamate (107). A typical bath contains 105 g L indium sulfamate to give about 30 g L of indium metal along with 150 g L sodium sulfamate, 26 g L sulfamic acid, 46 g L sodium chloride, and 2.3 g L triethanolamine. The bath is operated at pH 1.5–2.0 at room temperature with 100–200 A m²; anodes are indium metal. Higher indium content allows current up to 1000 A m². The metal is expensive, and plating is not widely practiced.

Iron. Iron plating is not widely used. However, copper alloy soldering iron tips are plated with thick (175–200 μm) iron plating to prevent solder from dissolving the copper. Iron has also been used to plate stereotype printing plants, and in plating of aluminum pistons and cylinders in automotive engine blocks. Mismachined and worn parts have been salvaged on a small scale with iron plating. Some sleeves used on larger engines have been iron plated. Printing plates for printing currency have been electroformed in iron plating baths. Iron has been plated from chloride, sulfate, mixed chloride–sulfate, fluoborate, and sulfamate solutions; chloride is the most common.

Lead and Lead–Tin Alloys. Lead is plated from fluoborate solutions, as shown in Table 13. Lead is used in battery parts for its good resistance to sulfuric acid (see BATTERIES). Low tin alloys of lead are used in bearings; 93° o lead–7° o tin was specified for bearings on piston aircraft engines. Lead–tin alloys of 3–15° o tin are calledterne; for hot-dip coatings, temeplate can be 20° o tin. Terneplate has been used to protect steel in gasoline tanks. Solder deposits are also plated. Lead–tin baths are similar to lead baths. Stannous fluoborate is added to supply the tin. Sulfamate solutions have been described for lead plating (108,109), but have found little industrial use. Requirements for lead and lead–tin deposits on steel are specified (110). These deposits have outstanding resistance to atmospheric corrosion, especially with a 0.25 μm copper strike undercoat. Lead coatings (19 μm) show an expected life of more than nine years in an industrial atmosphere. Up to 15° o tin does not decrease the corrosion resistance. Newer baths based on methane sulfonic acid are gaining in use both for lead (qv) and lead–alloy (see LEAD ALLOYS) baths. Lead is under strict regulation in the environment and waste streams. The future of lead plating is uncertain.

Table 13. Lead and Lead–Tin-Plating Baths

Parameter	Lead	7° o Sn–Pb	Solder, 60° o Sn–Pb
lead metal, g L	120	88	30
Pb(BF ₄) ₂ , ^a g L	220	160	53
HBF ₄ free, g L	30	100–200	100–200
H ₃ BO ₃ , free, g L	13.3		30
Sn metal, g L	6	52	
Sn(BF ₄) ₂ , ^a g L		15	130
peptone, ^b g L	0.2–1.0 ^c	0.5	5.0
temperature °C	25–40	15–38 ^d	15–38
current density, A m ²			
cathode	50–500 ^e	300	300
anode	100–300	150	

^a Available commercially as liquid concentrates.
^b Used as brightener.
^c Animal glue or gelatin has also been used, 0.2 g L. Peptone is predissovled and stored warm for a few days before use.

^d Higher temperatures decrease Sn in deposit.
^e Doubling the concentration of the chemicals, except the brightener, permits higher currents, up to 700 A m² for barrel plating and heavy buildup.

Nickel. Nickel plating continues to be very important. Many plating baths have been formulated, but most of the nickel plating is done in either Watts baths or sulfamate baths. Watts baths contain sulfate and chloride nickel salts along with boric acid, and were first proposed in 1916 (111). Nickel was first plated from sulfamate in 1938 (112) and patented in 1943 (108). The process was brought to market in 1950 (113). Typical bath compositions and conditions are shown in Table 14.

Table 14. Nickel-Plating Baths

Parameter	Watts	Watts high chloride	Sulfamate
Ni metal, ^a g L	82	77	75
NiSO ₄ ·6H ₂ O, ^b g L	300	135 ^c	
NiCl ₂ ·6H ₂ O, ^b g L	60	190	^d
Ni(SO ₃ NH ₂) ₂ ·4H ₂ O, ^b g L			410
H ₃ BO ₃ , ^e g L	35–45	35–45	35–45
pH	2–4	2–4	3.5–4.5 ^f
temperature, °C ^g	50–60	50–60	40–60
cathode current density, ^h A m ²	up to 430	up to 500	up to 550
anodes	Ni, electrolytic S-bearing	electrolytic Ni	S-bearing Ni ⁱ

^a As Ni is increased, the H₃BO₃ should be reduced.
^b Nickel salts are commercially available as liquid concentrates.
^c Proportions of Cl⁻ and SO₂₄ can be varied; total metal is usually maintained at 70–90 g L.
^d Cl⁻ is kept low in sulfamate, concentrations run from zero to about 3 g L as Cl⁻; NiCl₂·6H₂O, MgCl₂·6H₂O, and KCl have been used.
^e H₃BO₃ is kept at near-saturation; solubility varies directly with temperature.
^f Higher temperatures and lower pH should be avoided.
^g Temperature should be kept constant to maintain H₃BO₃ level. Low (45°C) temperature Watts baths can be run if Ni is reduced to about 45 g L total and H₃BO₃ is increased to about 50 g L.
^h Higher currents are attained using higher agitation, higher Ni, and higher temperatures.
ⁱ If using other anodes, sulfamate ion can be oxidized.

Both Watts and sulfamate baths are used for engineering application. The principal difference in the deposits is in the much lower internal stress obtained, without additives, from the sulfamate solution. Tensile stress can be reduced through zero to a high compressive stress with the addition of proprietary sulfur-bearing organic chemicals which may also contain saccharin or the sodium salt of naphthalene-1,3,6-trisulfonic acid. These materials can be very effective in small amounts, and difficult to remove if overadded, eg, about 100 mg L of saccharin reduced stress of a Watts bath from 240 MPa (34,800 psi) tensile to about 10 MPa (1450 psi) compressive. Internal stress value vary with many factors (22,71) and numbers should only be compared when derived under the same conditions.

Potential problems when additives are used in nickel-plating baths include co-deposition of the sulfur compounds. Sulfur-bearing stress reducers can increase brightness, brittleness, strength, and hardness, and decrease corrosion resistance, heat resistance, and resistance to oxidation. For applications that require low stress and pure deposits, sulfamate baths are used. In a newer two-stage method of electroforming large nickel textile printing screens, Watts baths with a stress reducer in the first stage and a brightener stress reducer in the second stage are used to make high quality screens in high volumes (114).

Sulfamate nickel has some other differences in operation. First, under conditions of higher temperatures and lower pH, the sulfamate ion can hydrolyze to form sulfate and ammonium ion. High ammonium ion increases stress and decreases ductility at higher currents. Ammonium ion cannot be removed practically from the plating bath. The second problem concerns a phenomenon where the sulfamate ion is oxidized at an anode surface that is not completely active. Higher chloride content is helpful when nickel anodes having a slight passivity are employed. Completely inactive or insoluble anodes can oxidize the sulfamate rapidly. The oxidized sulfamate acts much like the sulfite or dithionite ion in that some sulfur ion is co-deposited, and the deposit has properties similar to those obtained using saccharin. As little as 0.26 A·h L with carbon anodes drove the stress from 41 MPa (5945 psi) tensile to 131 MPa (19,000 psi) compressive when tested at 15 μm thickness at 270 A m² using the spiral contractometer (22) and ASTM B636. In a similar test using electrolytic nickel at a low anode current density, 1.7 A·h L drove the stress from 21 MPa (3045 psi) tensile to 48 MPa (6960 psi) compressive. The principle of using a portion of the tank current to go through an inactive, or less than 100⁰ % active, anode as a means of stress control has been promoted (115). If conditions require the use of insoluble auxiliary anodes, higher chloride seems to retard the oxidation. If the stress becomes too compressive, an electrolytic recovery process (22) has been used. The use of active sulfur-containing nickel anodes has eliminated the anode oxidation problem.

Nickel sulfamate is more soluble than the sulfate salt, and baths can be operated using higher nickel concentrations and higher currents. Sulfamate baths have been found to have superior microthrowing power, the ability to deposit in small cracks or crevices. Using one nickel salt, only a hydrometer and pH paper are needed to control the bath. A small amount of chloride salt was added as a proprietary. Highly purified nickel sulfamate concentrates are commercially available that can be used to make up new plating baths without further purification.

For Watts baths, nickel sulfate provides most of the nickel ion, (see NICKEL COMPOUNDS). Nickel chloride aids in anode corrosion, conductivity, and throwing power, but increases tensile stress. Boric acid is a buffer that works principally in the cathode film. It reduces burning and pitting, allows the use of higher currents, and contributes to smoother, more ductile deposits. Common impurities that affect nickel baths include iron, copper, zinc, chromium, and lead. Treatments and some additives for removal or control of these have been developed so that nickel plating baths can be used for many years with proper attention. When making up a new bath, it is often necessary to purify the solution before it can be used. Typically, high pH, 5.0–5.2, and activated carbon treatments are used along with several hours of low current density (LCD) electrolysis. Iron is removed by high pH precipitation, or sometimes complexed with citrates or gluconates. Anode efficiency is higher than cathode efficiency, and using closed loop systems, nickel ion builds up and eventually solution has to be withdrawn. Several methods are available for recovery of nickel from waste streams, or it is easily precipitated as the hydroxide sludge. Costs are becoming more favorable to recover the nickel.

Wetting agents such as sodium lauryl sulfate are usually necessary to reduce pitting in most nickel baths. Improved proprietary wetting agents are available with less foaming and better antipit properties. One nickel process that rarely needs wetting agent is barrel plating. Nickel plating in a barrel with the sulfamate bath often requires higher (5–6 g L) chloride, temperatures of 60–63°C, and lower (3.3–3.6) pH, to avoid laminated plate.

Nickel strike solutions are used primarily to obtain good adhesion on nickel, stainless steels, and other passive substrates. The most common is

called Woods strike, composed of nickel chloride and hydrochloric acid in various proportions. One common formula contains 240 g L nickel chloride hexahydrate and 12° by volume hydrochloric acid. Nickel glycolate strikes have had limited use for plating directly on zinc-based die castings. All sulfate, sulfamate, and Watts nickel strikes have been used in plating on plastics. A low (1.5 or less) pH nickel sulfamate strike, operated at room temperature, has been used for many years for plating on multimetal composites and passive metal substrates. Electroplated nickel coatings for engineering uses are covered by ASTM B689 (33).

The physical and mechanical properties of electroplated nickel vary with the type of bath, impurities, and additives used, along with operating conditions such as pH, current density, temperature, and agitation. Plating thickness should always be included when expressing values for elongation and other mechanical properties. Generally, nickel deposits from sulfamate baths are a little harder and stronger, with less elongation than Watts deposits (116).

Bright Nickel. Nickel as electroplated from pure nickel salts is a dull grey, satin-like deposit that has to be buffed to obtain a bright finish. Brighteners are used to obtain bright deposits directly from the bath (117). The additives currently used fall into two classes, which have variously been labeled primary and secondary, first class and second class, and carrier and brightener. The last is more commonly used in plating plants.

The carrier portion of the system contains materials that have been found to produce some grain-refinement, some brightness, and contribute to more compressively stressed deposits. The brightener portion, sometimes called brightener-leveler, contains materials that in conjunction with the carrier produce high luster and leveling to build considerable brightness. These materials often contribute to high tensile stress. Impurities have to be kept low and plating solutions well-filtered. Bright nickel baths should be periodically purified using low current density (LCD) electrolysis and activated carbon treatments. Wetting agents are usually required to control pitting.

The additives that go into the nickel bath are almost always proprietary; carrier portions may include the sodium salts of benzene disulfonic acid, naphthalene 1,5-disulfonic acid, saccharin, and allyl sulfonic acid. Generally, these compounds are characterized by a $\text{—C(SO}_2\text{—)}$ group. Brightener-leveler portions include a large number of possible materials. A few are based on small amounts of zinc, cadmium, lead, selenium, or tellurium, etc, and many are based on organic compounds containing unsaturated groups such as C=O , C=C , $\text{C}\equiv\text{C}$, C=N , $\text{C}\equiv\text{N}$. Nickel brighteners and the theories of brightening are well-covered in patents and other literature (118). When using carrier-brightener additives, portions of these materials are codeposited; as much as 0.06 to 0.10° sulfur, as sulfide, or higher is present. High sulfur content increases the activity of the nickel deposit. This is not always a desirable case. Bright nickel baths are usually based on Watts bath formulas, regular or high chloride. Sulfamate solutions have been used, but on a much smaller scale.

Dual Nickel. Semibright nickel is a term given to sulfur-free systems. The more well-known of these were based on coumarin (qv). At first, these soft, hazy deposits were buffed bright and a layer of bright nickel deposited for appearance. It was later observed that better corrosion protection was obtained using this dual nickel system; not all of which could be attributed to the buffing. This improvement is ascribed to the sulfur-free nickel being much more noble than the more anodic top layer; pit penetration as it reaches the cathodic layer is reduced and directed more laterally. More recently this potential difference is being used in specifications as the STEP test (119). The sulfur content in the semibright layer should be less than 0.005 mass ° and more than 0.04 mass ° in the bright nickel. Improvements in sulfur-free additives have resulted in much brighter deposits and the term semibright could be replaced by sulfur-free.

Triple Nickel and More. As an extension to the dual nickel, a thin, higher sulfur-containing nickel strike is deposited between the sulfur-free and the bright nickel. The sulfur content of this minimally 2.5 μm -strike is 0.15–0.20 mass °. Triple nickel and dual nickel are covered by ASTM specification B456 (89). A fourth nickel deposit has shown improved protection by the effects it has on subsequent chromium deposits. Highly stressed, these nickel strikes have been used to aid in producing microcracked chromium.

Special Nickel-Plating Solutions. Black nickel for nonreflective, decorative, or solar absorptive uses is plated from nickel solutions containing ammonium, zinc, and thiocyanate salts. Very low (5–20 A m²) currents are used (120); coatings are thin and often waxed to protect the finish. Nickel composites containing hard particles for wear, or abrasive particles for cutting and grinding tools, are plated from conventional Watts or sulfamate solutions to which the particles are added. Special techniques are required to obtain even particle deposition throughout the deposit. Other composites having plastic powders, fibers, or PTFE for lubricity require additives to aid in wetting and dispersion.

Nickel Alloys. The electrodeposition of nickel alloys and other metal alloys is discussed in detail in Reference 121.

Nickel–Iron. Nickel alloys containing 10–35° iron are used in decorative applications as a cost-saving substitute for all-nickel deposits. One Watts-based variation contains about 45 g L nickel, 3 g L iron, and boric acid and additives such as citrates and gluconates. Hydroxy carboxylic acids act as stabilizers for the ferrous iron. The bath is run at 55–65°C at a pH of 2.8–3.6 and currents of 300–400 A m². Iron salts are added as ferrous sulfate or ferrous chloride. Bright nickel additives are used. These baths reached a peak in use in the late 1970s and have since decreased. A small amount of nickel–iron having 21° iron, often called permalloy for the high permeability, is plated for its magnetic properties.

Nickel–Cobalt. Alloys of nickel–cobalt (122), are used for engineering properties, especially in electroforming. Deposits having 1–50° cobalt have been obtained from various baths. Nickel–cobalt plating solutions are similar to those used for engineering nickel with cobalt salts substituted for some of the nickel salts. Sulfate, chloride, and sulfamate salts of cobalt are used. Cobalt adds strength and hardness to nickel deposits. Cobalt content of the deposit increases with bath agitation and with the ratio of cobalt to nickel in solution. Generally, the cobalt in the deposit decreases with current density; however, many contradictory effects on this and other plating solution variables have been reported (123).

Nickel–Phosphorus. Interest in electrodeposited nickel–phosphorus came with realization of the benefits of the electroless nickel-plated alloy. The properties of the alloys appear to be the same from either process and are related to the phosphorus content (124). Deposits using 2–15° phosphorus and higher can be electroplated; lower phosphorus deposits are dull and higher phosphorus deposits are bright. Phosphite ion provides the phosphorus; the more expensive hypophosphite used in electroless nickel is not required. Low (2°) phosphorus deposits, can be obtained simply using phosphorus acid additions to a Watts bath and operating the bath at 0–2 pH with a temperature of 75–95°C and current of 500–4000 A m². Higher phosphorus alloys require higher phosphite in solution. To obtain this without making the bath too acidic, part of the nickel content is added as the reaction product of nickel carbonate and phosphorus acid. Some nickel phosphorus alloy is being electroplated on large shafts that used to be electrolessly plated.

Nickel–Tungsten. Alloys can be deposited from acid Watts baths that contain sodium tungstate producing deposits with up to about 4° tungsten. Alkaline baths having ammonium salts in the nickel solution, along with Rochelle salts, citrates, or glycolates can deposit 10–20° tungsten. No significant commercial uses of nickel–tungsten alloys have been publicized, although the properties may be of interest in special applications (125). Thin nickel–tungsten coatings having less than 4° tungsten have been found to solder very well even after aging (126). Nickel–tungsten alloys having codeposited silicon carbide particles (127) have been proposed as a possible replacement for hard chromium plating. The deposit contains 5–10° silicon carbide and 44–50° tungsten.

Platinum-Group Metals.

Rhodium. Rhodium is the most commonly plated platinum-group metal. In addition to its decorative uses, rhodium has useful properties for engineering applications. It has good corrosion resistance, stable electrical contact resistance, wear resistance, heat resistance, and good reflectivity. The use of rhodium for engineering purposes is covered by an ASTM specification (128). Typical formulas are shown in Table 15. The metal content is obtained from prepared solutions available from proprietary plating supply companies. Replenishment is required because anodes are not soluble. Rhodium for decorative use may be 0.05–0.13 μm thick; for industrial use, it may be 0.50–5.0 μm thick.

Table 15. Rhodium-Plating Baths^a

Parameter	Decorative		
	Phosphate bath	Sulfate bath	Industrial
rhodium, g L ^b	2	1.3–2.0	5
phosphoric acid (85°), mL L	40–80		
sulfuric acid (96°), mL L		25–80	25–50

current density, A m ²	200–1000	200–1000	100–300
temperature, °C	45–50	45–50	45–50

^a Anodes can be platinum or platinum-clad.
^b Rhodium from commercial concentrate in appropriate acid, phosphoric or sulfuric.

Platinum. Platinum plating has found application in the production of platinized titanium, niobium, or tantalum anodes which are used as insoluble anodes in many other plating solutions (see METAL ANODES). Plating solutions were often based on platinum "P" salt, which is diamminedinitroplatinum(II). A dinitroplatinite sulfate–sulfuric acid bath has been used to plate directly onto titanium (129). This bath contains 5 g L of the platinum salt, pH adjusted to 2.0 with sulfuric acid. The bath is operated at 40°C at 10–100 A m². Other baths based on chloroplatinic acid have been used in both acid and alkaline formulations: the acid bath uses 20 g L of the platinum salt and 300 g L hydrochloric acid at 65° C and 10–200 A m². The alkaline bath uses 10 g L of the platinum salt, 60 g L of ammonium phosphate and ammonium hydroxide to give a pH of 2.5–9.0. The alkaline bath can be plated directly onto nickel-base alloys; acid baths require a gold strike on most metals.

Palladium and Palladium Alloys. Palladium is used in telephone equipment and in electronics applications as a substitute for gold in specific areas. Palladium is plated from ammoniacal and acid baths; available along with chelated variations as proprietary processes. One typical alkaline bath uses 8 g L diammine-dinitropalladium, 100 g L ammonium nitrate, and 10 g L sodium nitrite. The pH is adjusted to 9–10 using ammonium hydroxide, and the bath is operated at 100 A m² at 50° C . If ammonium sulfamate, 100 g L, is used in some baths to replace the nitrate and sodium nitrite salts, the bath is run at lower temperature, 25–35°C, and a pH of 7.5–8.5. A palladium–nickel alloy, 75° Pd, is plated from a bath having 6 g L palladium from the same salt, 3 g L nickel from nickel sulfamate concentrate, and 90 g L ammonium hydroxide. The bath is operated at 20–40°C with 50–100 A m².

Other Metals. Ruthenium, the least expensive of the platinum group, is the second best electrical conductor, has the hardest deposit, and has a high melting point. A general purpose bath uses 5.3 g L of ruthenium as the sulfamate salt with 8 g L sulfamic acid, and is operated at 25–60°C with a pH of 1–2. Osmium has been plated from acid chloride solutions (130) and iridium from bromide solutions, but there are no known applications for these baths.

Silver Plating. Cyanide baths are the standard for silver plating. Only minor modifications and improvements in brighteners have occurred since silver electroplating was first patented in 1840 (16). Typical plating baths are shown in Table 16. Silver strikes are necessary to obtain adhesion on most metals. Although a large proportion of silver plating is used for decorative purposes, such as tableware, jewelry, art work, musical instruments and the like, a growing amount is found in industrial applications. These engineering uses include: bearings, reflective surfaces, contact areas on busbar and other electrical and electronic contacts, solderable surfaces, in thermocompression bonding, selective plating on semiconductor components to replace gold on many devices (131), and other applications where the high thermal and good electrical conductivity are useful. There is an ASTM specification covering silver plating for engineering uses (132). Chromate conversion coatings are sometimes specified as a post treatment in engineering applications. A proprietary, noncyanide, slightly alkaline silver-plating bath based on succinimide is being marketed, although the system has some limitations in comparison with cyanide.

Table 16. Silver-Plating Baths^a

Parameter	Decorative	Industrial	Strike ^b
AgCN, g L ^c	30–35	45–50	1.5–5.0
KCN, g L ^d	50–78	65–72	75–90
free	35–50	45–50	
K ₂ CO ₃ , g L	15 ^e	40–80	
KNO ₃ , g L		40–60	
KOH, g L		1–14	
temperature, °C	20–28	42–45	22–30
current density, A m ²	50–150	500–1000	130–300

^a Anodes are normally pure silver, 999 +, often bagged.
^b For strike on nonferrous metals and a second strike on steel, first strike uses 1.5 g L AgCN, 75–90 g L KCN, and 10–15 g L copper cyanide.
^c Insoluble in water; dissolves in aqueous solution of KCN.
^d Some KCN is used to form the silver complex; the excess is called free CN . NaCN has been used, but has some deficiencies.
^e Value given in minimum.

A phenomenon known as silver migration has limited the use of silver plating in miniature circuit boards: under a positive d-c potential within a damp environment silver can migrate across insulation. On drying, silver is found in the insulation or media creating a leakage path. Silver plating is forbidden in many military specifications for circuit boards.

Tin Plating. Tin is plated from alkaline stannate and acid sulfate or fluoborate baths. (Table 17.) Deposits are matte when plated without brighteners. One older method of brightening is termed reflowing or flow-brightening where the plated tin part is heated for a few seconds just above the melting point and then quenched. Brighteners for use in sulfate baths have been vastly improved, and very bright, decorative deposits can be obtained directly from the bath. The key to operation of the stannate bath is in proper control of the anode current density to ensure that tin dissolves in the stannate state. If the proper anode film is not formed, stannite tin is generated with immediate poor results. Tin anodes containing about 1° aluminum increase the proper anode current density range.

Table 17. Tin-Plating Baths

Parameter	Stannate	Sulfate ^a	Fluoborate ^a
K ₂ Sn(OH) ^b , g L	100		
KOH, g L	15		
SnSO ₄ , g L		100	
H ₂ SO ₄ , g L		100	
Sn as Sn(BF ₄) ₂			40
HBF ₄ , free			40
temperature, °C	65–90	20–30	20–40
current density, A m ²	300–600	100–1000	200–1000
anodes	Sn, Fe or Sn–1° Al	Sn, Sn–1° Al	Sn, Sn–1° Al

^a Proprietary brighteners are used.

^b Sodium stannate and sodium hydroxide have been used, but at less efficiency, less conductivity, less current density, and more sludge.

Tin, as well as some other metals, can undergo a phenomenon where tiny metal filaments, called whiskers, form randomly on parts used in electrical applications. In low voltage, miniature circuitry, whiskers can cause short circuits. Alloys having 2% lead minimum or 0.5% bismuth or heat treatments are said to overcome the problem. A specification for electroplated tin coatings is available (133).

The newest tin baths, tin–lead baths, and lead baths recently entering the market, are based on methanesulfonates. The higher makeup cost of tin methanesulfonate baths, about 1.6 times the cost of fluoborate baths, may be justified where restrictions on fluoborates and boric acid in wastes exist. A special bath for high speed plating of strip steel using all halogens was introduced in 1942 (134).

Tin Alloys.

Tin–Lead. Tin–lead alloys are plated from fluoborate baths and more recently methanesulfonate solutions. Solder 60% tin plate is used on contacts and as a solderable etch resist on printed circuit boards. Higher (96–98% tin) tin alloys are used in semiconductor and other electronic applications. For 60% tin alloys, proprietary methane–sulfonate baths may contain 16 g L tin, 10 g L lead, 20% by volume methane–sulfonic acid, and the brightener. Solution is operated at 25°C and 100–400 A m². A fluoborate bath to produce 60% tin deposits contains about 60–65 g L tin, 25–26 g L lead, 40 g L fluoboric acid, and 23–30 g L boric acid along with a brightener; 6 g L of glue has been used. This bath is operated at 20–25°C up to about 320 A m². More acidic, dilute baths having 15 g L tin, 10 g L lead with 400 g L fluoboric acid, and 30 g L boric acid are used for high throw baths. Tin–lead alloys become lead–tin alloys when tin is less than 50%. Anode material for tin–lead plating baths should be of the same alloy as being plated.

Tin–Nickel. Alloy deposits having 65% tin have been commercially plated since about 1951 (135). The 65% tin alloy exhibits good resistance to chemical attack, staining, and atmospheric corrosion, especially when plated copper or bronze undercoats are used. This alloy has a low coefficient of friction. Deposits are solderable, hard (650–710 HV₅₀), act as etch resists, and find use in printed circuit boards, watch parts, and as a substitute for chromium in some applications. The rose-pink color of 65% tin is attractive. In marine exposure, tin–nickel is about equal to nickel–chromium deposits, but has been found to be superior in some industrial exposure sites. Chromium topcoats increase the protection further. Tin–nickel deposits are brittle and difficult to strip from steel. Temperature of deposits should be kept below 300°C.

The most common plating bath uses fluoride to complex the tin. A typical solution contains 45 g L stannous chloride, 300 g L nickel chloride hexahydrate, and 55 g L ammonium bifluoride. It is operated at pH 2.0–2.5 using ammonium hydroxide; temperature is 65–75°C and current about 200 A m². The bath has excellent throwing power. Air agitation is avoided. The deposit is bright without additives. Anodes are cast nickel, and the tin is replenished by additions of stannous chloride. Alloy anodes of 72% tin have been used to a much lesser extent. Tin–nickel deposits are covered by ASTM (136) and ISO (137) specifications. One other bath based on pyrophosphate has appeared in the literature, but does not seem to be in commercial use.

Tin–Zinc. Baths for tin–zinc alloys stem from work done after World War II in efforts to find a substitute for cadmium. Although alloys of all concentrations are possible, 80% tin–20% zinc gives the best combination of properties. This alloy has a low coefficient of friction, low electrical contact resistance, is solderable, slightly anodic to steel, and does not form voluminous corrosion products. In addition, the tin–zinc alloy has good paint adhesion qualities, good ductility, and is easily spotwelded.

A typical bath is based on stannate and cyanide; for 80% tin, the solution is made using 120 g L potassium stannate, 11.3 g L zinc cyanide, and 30 g L potassium cyanide. The bath is operated at 65°C with cathode current of 100–800 A m² and anode current of 150–250 A m². Anodes are the same composition as the alloy, and have to be filmed properly as for stannate tin plating.

Salt spray tests, humidity tests, and other accelerated tests, some using sulfur dioxide and carbon dioxide, have shown favorable results for tin–zinc in comparison with zinc, cadmium, and tin deposits. Chromating improves the performance.

A low tin alloy, more correctly 90% zinc–tin alloy, has been proposed as an economical solderable substitute for cadmium. For this 10% tin deposit, the same type of plating bath, but using 40 g L tin, 15 g L zinc, 15 g L potassium cyanide, and 45 g L free potassium hydroxide is used.

Other Tin Alloys. Proprietary tin–cobalt alloys having a third metal were used for a short time as a substitute for chromium in decorative applications. In spite of the cost, growing restrictions on chromium may bring renewed interest in tin–cobalt alloys.

Zinc. Zinc plating continues to be a popular, cost-effective way to protect steel from atmospheric corrosion. Very bright decorative, chromium-like coatings can be obtained. Post-plate treatments in chromate conversion coatings are used to extend corrosion protection and aid in paint adhesion. Chromate conversion coatings are available in several types, from the thinner blue-bright coatings, through the thicker yellow iridescent ones, to the thickest olive-drab coatings. Olive-drab chromates can extend salt spray protection of zinc coatings up to 400–500 hours. Fresh chromate coatings can be dyed a variety of colors, and the process is used for part identification.

Zinc is plated on pipe, conduit, strip, sheet, and wire, as well as sheet metal stampings, all sorts of hardware, and fasteners. Zinc plating is used on electronic housings and cabinetry to reduce high frequency radiation interferences.

Plating solutions for continuous strip, called electrogalvanizing, sheet, and wire are usually simple zinc sulfate solutions, although chloride and mixed variations have found some use. A typical bath for strip may contain 350 g L zinc sulfate heptahydrate and 30 g L ammonium sulfate. Baths are operated at pH 3.0–4.5, 40–55°C, at currents of 1000–6000 A m². Zinc thickness is often 0.3–4.3 μm. A wide variety of additives have been used to produce some grain refinement, eg, dextrin, licorice, glucose, and gelatin. Wire plating, which may have 13–130 μm zinc plate, has been plated in more acidic zinc sulfate solutions using insoluble silver–lead alloy anodes. Currents up to 30,000 A m² are reported.

For general zinc plating in the United States, estimates of the zinc baths in use in 1970 showed cyanide at over 90%, chloride at 3%, and noncyanide alkaline (zincate) baths at 4%. By 1990, cyanide zinc was 20%, chloride 50%, and zincate 30% (138). In 1992 (139) the cyanide was 16%, chloride 48%, and zincate 36%. Moreover, the cyanide zinc baths of 1992 were more likely to be run at 15–30 g L sodium cyanide in contrast to the 100 g L normal in 1970. Each of the three zinc systems uses proprietary brighteners.

Chloride Baths. Chloride zincs are brighter, can plate directly on cast iron, and operate at high efficiency, but have poor throwing power, require acid-proof equipment, and are more corrosive to surrounding steel materials. A typical chloride zinc may contain about 35 g L zinc, 135 g L Cl⁻, 30 g L boric acid, and proprietary brighteners and wetting agents. The chloride source in earlier baths was the ammonium salt, which did not require boric acid. Potassium chloride is now more common, although some sodium chloride systems are used as well as mixtures with low ammonium salts. Higher currents are possible with the ammonium chloride baths, followed by potassium and then sodium. High purity zinc anodes, which are usually bagged to reduced roughness in rack plating, are used. Brighteners are slightly soluble organic materials solubilized in the plating solution with mixtures of wetting agents. Dilute baths with 15 g L zinc or less have been used in barrel plating of small parts such as nails. Baths require regular peroxide additions to precipitate iron, often on a daily basis; the ferric iron is then filtered out. Baths are operated at pH 4.8–5.4, with currents up to about 325 A m², and at ambient temperatures. More recently, baths have been operated warmer (44°C) to increase evaporation rate. This reduces tendency of zinc solution to increase in volume and flood the tanks. At low surface tension, drag-out is less than drag-in.

Zincate Baths. Zincates are simple, economical, noncorrosive to steel and have superior throwing power. Newer systems are showing better tolerance to operating variables, with less tendency toward flaking or delayed blistering, and good brightness. Zincate baths, however, require good cleaning, frequent analysis, close chemical control, and have low cathode efficiency. Zinc plating is discussed in detail in the literature (81).

Alkaline zincate bath formulations vary according to the brighteners used. Zinc at 6–12 g L is common; newer baths have zinc at 9–18 g L. Sodium hydroxide is about 10 times the zinc, and is varied to control the rate of zinc dissolution. Sodium carbonate, about 30 g L, is added to new baths for conductivity and to reduce impurity effects. Zinc anodes corrode when the bath is idle. To resolve this and anode-caused roughness, only steel anodes are used in the plating tank in many plants. The zinc is replenished electrochemically through a separate generator tank where plating solution is recirculated through zinc balls or slab in steel baskets. Deposits from the zincate are powdery and noncoherent without additives. Additives, needed for both adhesion and cohesion, are almost always proprietary, and contain mixtures of organic brightening compounds along with certain polyamines or other materials that

perform the same functions as cyanide (140). Complexing agents are generally avoided because of possible effects on waste treatment and codeposition of metallic impurities. Zincate baths are operated at 27–40°C at currents up to 300 A m².

Cyanide Baths. Cyanide zincs have excellent covering power and throwing power, good cleaning ability, deposit relatively pure zinc, are capable of thick deposits, and require little control. Zinc deposits from cyanide baths are purer than from the other baths, but still contain traces of the metal contaminants, brightener components, sulfur, hydrogen, and other gases. Pure, sulfur-free zinc is resistant to hydrochloric acid (141,142).

Zinc coatings are covered by an ASTM specification (143). Because of the varying purity of zinc deposits from chloride and zincate baths, some thickness measuring methods can vary considerably (39). Methods based on physical measurement of magnetic methods (144,145) are used for best accuracy.

Other Baths. Other forms of zinc plating are also in use. Immersion zinc deposits are used as a preparatory step in electroless plating or electroplating of aluminum (146), magnesium (147), and beryllium (148) alloys. Formulations vary with the application; typical baths are listed in the references cited.

A mechanical plating method, often called peen plating, uses zinc dust, tumbling equipment, glass beads, or other impact media, along with some proprietary chemicals. No current is used. An advantage is that hydrogen embrittlement is avoided, and springs and other hardened steel parts can be plated safely (149).

Zinc Alloys. There has been considerable worldwide activity in the area of plating zinc alloys. This interest results from efforts to improve the corrosion resistance of automobiles and automotive components without using cadmium (150). The use of zinc–alloy plated coil for automotive body steel originated in Japan, has spread through Europe, and more recently has come to the United States. Three zinc alloys dominate the interest: zinc–nickel, zinc–iron, and zinc–cobalt. Europe produces predominantly zinc–nickel. In Japan, zinc–nickel and zinc–iron are more popular (151). The annual worldwide nickel consumption for zinc–nickel alloy plating has been estimated at 2700–3175 metric tons (152). In the United States, consumption is estimated at about 225 metric tons for this purpose. Alloys are generally 6–12% nickel. Usage is expected to increase.

As of this writing the zinc alloys are too new to have actual corrosion resistance data, except for that based on accelerated tests. Zinc–nickel usually shows better results than zinc–cobalt in salt spray tests. The reverse is true when the Kesternich test is used. Tin–zinc performs well in both salt spray and Kesternich tests, but appears only to equal zinc plating and zinc–nickel in humidity tests.

Zinc–Nickel. Steel has the best salt spray resistance when the nickel is 12–13% of the alloy. At increasing nickel contents, the deposit becomes more difficult to chromate and more noble, eventually becoming cathodic to steel. At those levels and above, corrosion resistance usually decreases and is dependent on a complete lack of porosity for protection of the steel. In efforts to replace cadmium and nickel–cadmium diffused coatings in the aircraft industry, zinc–nickel has insufficient wear properties for some application, but is under study as an undercoat to various electroless nickel top coats (153).

When zinc–nickel plate is to be painted, whether on strip or on component parts, plate thickness of 4 μm or more is effective in reducing corrosion creepage on scribed test panels. Minimum creep is at about 13 μm (151). Typical bath formulas for zinc–nickel and the other zinc alloys are given in Table 18. Acid zinc–nickel baths used on strip lines may not use additives. All other zinc–alloy baths use brighteners and other additives, almost always proprietary. The chloride or alkaline bath options behave similarly to all-zinc baths with regards to throwing power and efficiencies. Deposits from acid zinc–nickel give higher nickel deposits, but the alloy varies considerably with current density among other variables. Deposits from alkaline baths are lower in nickel but more consistent in both nickel content and plate thickness distribution. Anodes for acid zinc–nickel are usually zinc along with some graphite anodes; the proportion is adjusted to prevent excessive zinc concentrations. Anodes for the alkaline system are usually steel and the zinc is supplied by dissolving zinc anodes in a separate generator tank. The alloying metal is added as the metal salt to maintain a set ratio. Zinc–nickel alloys, as well as the other zinc alloys, vary widely with changes in the system; processes need good control of the chemical constituents including additives along with temperature, agitation, and current density, to produce consistent alloys. Corrosion resistance benefits for lower nickel contents have been reported on fasteners with 2–3% nickel, and in another process with nickel contents as low as 0.25–0.75% nickel (154). High salt spray protection is also claimed using a duplex coating with higher nickel undercoating with a lower nickel alloy topcoat (155).

Table 18. Plating Baths for Zinc Alloys

Parameter	Zn–Co		Zn–Ni		Zn–Fe
	Acidic ^a	Alkaline	Acidic ^{b,c}	Alkaline	
zinc, g L	25	8	30	8	8
cobalt, g L	4	0.04			
nickel, g L			25	1.6	
iron, g L					0.05
total chloride, g L	135		240		
boric acid, g L	25				
sodium hydroxide, g L		90		130	90
temperature, °C	29	25	40	22	22
cathode current density, A m ²	215	320	320	320–850	320
deposit, %	0.3–0.8% Co		10–15% Ni		0.3–0.8% Fe
anodes	zinc	steel	zinc	Ni-plated steel	steel

^a pH 5.0.
^b pH 5.8.
^c Also contains 4% NH₄OH by volume.

Zinc–Cobalt. Alloys of Zn–Co usually contain 0.3–0.8% cobalt. Higher cobalt alloys, from 4–8%, have shown better salt spray resistance (156), but the commonly plated alloy is 0.3–0.8%. One automotive company specifies 0.3–1.0%. Cobalt is expensive, and economics favor the lower alloys. Costs have been quoted for zinc–cobalt at 1.2 times the cost of chloride zinc, with zinc–nickel alloys at 1.5–1.6 times the chloride zinc. Deposits can be very bright, but the improved corrosion resistance advantage requires yellow or bronze chromates. Alkaline baths give fewer problems in plating components with lapped, spot-welded seams.

Zinc–Iron. The Zn–Fe alloy is plated from an alkaline bath. Deposits are 0.3–0.8% iron and can be given attractive, resistant, black, silver-free chromate coatings. Corrosion protection requires the heavier, darker chromates. Zinc–iron baths are the most economical of the zinc alloys.

Post-Treatments. Although many post-treatments have been used over plated metals, chromate conversion coatings remain as the most popular. Chromates are used to improve corrosion resistance, provide good paint and adhesive base properties, or to produce brighter or colored finishes. Formulations are usually proprietary, and variations are marketed for use on zinc, zinc alloys, cadmium, copper and copper alloys, and silver (157). Chromates are also used on aluminum and magnesium alloys (158,159). More recently, chromate passivation has been used to extend salt spray resistance of autocatalytic nickel plated parts.

For zinc deposits, various chromates can produce very thin, blue-bright coatings to get up to about 24 hours salt spray protection before the zinc corrodes (white corrosion salts), thicker yellow iridescent films for up to about 100 hours, and thickest, olive-drab coatings that may give upward of 300 hours (160). A modification of the olive-drab chromate solution with a silver salt produces a black finish, but salt spray resistance drops back to 48–72 hours. Topcoats of waxes are often used on such blacks. Higher salt spray resistance is being claimed for chromates on the zinc alloys; eg, 250 hours for

yellow chromate on zinc–cobalt, 1000 hours for yellow on zinc–nickel, and 300 hours on a black over zinc–iron electroplate. This black chromate does not contain silver (161). The higher nickel contents in zinc–nickel alloys are difficult to chromate; at this time no olive-drab or black coatings are marketed.

Electroforming

Electroforming is the production or reproduction of articles by electrodeposition upon a mandrel or mold that is subsequently separated from the deposit. The separated electrodeposit becomes the manufactured article. Of all the metals, copper and nickel are most widely used in electroforming. Mandrels are of two types: permanent or expendable. Permanent mandrels are treated in a variety of ways to passivate the surface so that the deposit has very little or no adhesion to the mandrel, and separation is easily accomplished without damaging the mandrel. Expendable mandrels are used where the shape of the electroform would prohibit removal of the mandrel without damage. Low melting alloys, metals that can be chemically dissolved without attack on the electroform, plastics that can be dissolved in solvents, are typical examples.

The surface of the mandrel is reproduced exactly down to submicroscopic detail on the electroform. This duplication of detail accounts for many applications for electroforming. The U.S. Bureau of Engraving uses about 40 t yr of nickel in electroforming printing plates for currency and stamps. The principal use of copper in electroforming is in the continuous production of copper foil used in the printed circuit industry. Applications are much more diverse for nickel. The largest nickel application is in the manufacture of textile printing screens, where mandrels are rolls coated with a developed photoresist pattern, or a mandrel prepared by etching and masking to produce a screen or mesh pattern. Printing screens can be 6 meters long, over 30 cm in diameter, and 100–125 μm thick.

Nickel electroforms were used in making phonograph records, which are now quite rare; however, one of the replacements, the compact (audio) disk, utilizes a nickel electroform as a mold or stamper to produce these as well as video disks. More recently, computer disks for permanent data storage are being produced with an electroformed stamper. Applications are found from very thin filters of about 25 μm thick having microscopic holes, to large electroformed rocket motor components used in the U.S. space shuttle, which can be about 2-cm thick in some areas. Seamless cylinders, used in copying machines, have been produced in large volume on automatic plating machines. Many parts are manufactured by electroforming that would be difficult or impossible to make by other methods.

Plating solutions used in nickel electroforming are primarily the Watts bath and the nickel sulfamate bath. Watts baths exhibit higher stress and require additives for stress control, which may affect other properties. Sulfamate baths produce much lower stress and are preferred where purer nickel or nickel–cobalt deposits are needed. ASTM specifications are available that describe the mandrels and plating solutions (116,162).

BIBLIOGRAPHY

"Electroplating" in *ECT*1st ed., Vol. 5, pp. 611–646, by C. L. Faust and W. H. Safranek, Battelle Memorial Institute; in *ECT* 2nd ed., Vol. 8, pp. 36–74, by F.A. Lowenheim, M&T Chemicals, Inc.; in *ECT* 3rd ed., Vol. 8, pp. 826–869, by F.A. Lowenheim, Consultant.

1. *ASTM B374, Standard Terminology Relating to Electroplating*, American Society for Testing and Materials, Philadelphia, Pa., 1993.
2. W. Blum and G. Hogaboom, *Principles of Electroplating and Electroforming*, McGrawHill Book Co., Inc., New York, 1924.
3. W. Goldie, *Metallic Coating of Plastics*, Vols. 1 and 2, Electrochemical Publications Ltd., Ayr., Scotland, 1968–1969.
4. G. Pohl, ed., *Products Finishing Mag.*, private communication, 1992.
5. *Metal Finishing Guidebook and Directory*, '92, Metal Finishing, Hackensack, N.J., p. 853.
6. B.A. Graves, *Products Finishing* **57**(7) 66–75 (Apr. 1993).
7. G. Kraft, *Metal Finishing*, 29–31 (July 1990).
8. S. Etris, Gold Institute, Washington, D.C., private communication, 1992.
9. G. Dibari, INCO, Toronto, Canada, private communication, 1992.
10. *Electroless Nickel Conference*, Annecy, France, Apr. 3, 1992 *La Journée du Nickel Chimique*, Saddlebrook, N.J., 1992.
11. Technical data, Tin Information Center, Columbus, Ohio, 1992.
12. W. Hampshire, Tin Information Center, Columbus, Ohio, private communication, 1992.
13. G. Shawahan, Enthone OMI, New Haven, Conn., private communication, 1992.
14. G. DiBari in Ref. 5, p. 227.
15. **U.S. Pat. 2,678,909** (May 18, 1954), G. Jernsedy and J. Patrick (to Westinghouse Electric Corp.).
16. **Brit. Pat. 8447** (1840), G. Elkington and H. Elkington (to Elkington Co.).
17. D. Foulke and M. Diggin, in F. Lowenheim, eds, *Modern Electroplating*, 2nd ed., John Wiley & Sons, Inc., New York, 1963, p. 327.
18. J. Norris, in Ref. 5, pp. 301–313.
19. A. Vagramyan and Z. Soloveva, *Technology of Electrodeposition*, Robert Draper Ltd., Teddington, UK, 1961.
20. H. Haring and W. Blum, *Trans. Amer. Electrochem. Soc.* **44**, 313 (1923).
21. **U.S. Pat. 3,300,396** (Jan. 24, 1967), C. Walker (to Univertical Corp.).
22. J. Marti, *Plating* **53**(1), 61–71 (Jan. 1966).
23. C. Smith, in Ref. 5, p. 656.
24. *Surface Cleaning, Finishing, and Coating*, Vol. 5, 9th ed. ASM, Materials Park, Ohio, 1982.
25. W. Innes, in Ref. 5.
26. H. Silman, G. Isserdis, and A. Averill, *Protective and Decorative Coatings for Metals*, Finishing Publications, Ltd., Teddington, UK, 1978, Chapt. 3.2.
27. S. Spring, *Industrial Cleaning*, Prism Press, Melbourne, Australia, 1974.
28. J. Koelsch, *Manuf. Eng.*, 75–78 (Mar. 1992).
29. *ASTM B689, Standard Specification for Electroplating Engineering Nickel Coatings*, American Society for Testing and Materials, Philadelphia, Pa., 1990.
30. *ASTM B650, Standard Specification for Engineering Chromium Coatings on Ferrous Substrates*, American Society for Testing and Materials, Philadelphia, Pa., 1985.
31. *ASTM B656, Standard Guide for Autocatalytic Nickel–Phosphorus Deposition on Metals for Engineering Use*, American Society for Testing and Materials, Philadelphia, Pa., 1991.
32. *McKutcheon's 1992, Emulsifiers and Detergents*, Vol. 1, Manufacturing Confectioner Pub. Co., Glen Rock, N.J.
33. Ref. 5, p. 194.
34. Ref. 5, p. 197.
35. Ref. 5, p. 129.
36. C. Zapffe and M. Haslim, *Plating* **36**, 906 (1949).
37. "Metallic and Inorganic Coatings," *ASTM Annual Book of Standards*, Vol. 02.05, 1992. Issued annually; note that volume numbers are subject to change.
38. *ASTM B659, Std. Guide for Measuring Thickness of Metallic and Inorganic Coatings*, American Society for Testing and Materials, Philadelphia, Pa., 1990.
39. J. Homer, *Metal Finishing*, 15 (Oct. 1987).
40. *ASTM B117, Std. Test Method of Salt Spray (Fog) Testing*, American Society for Testing and Materials, Philadelphia, Pa., 1990.
41. J. Homer, *Metal Finishing*, (Feb. 1990).
42. *ASTM B368, Std. Method for Copper-Accelerated Acetic Acid Salt Spray (Fog) Testing (CASS Test)*, American Society for Testing and Materials, Philadelphia, Pa., 1990.

43. *ASTM B380, Std. Method of Corrosion Testing of Decorative Electrodeposited Coatings by the Corrodokote Procedure*, American Society for Testing and Materials, Philadelphia, Pa., 1990.

44. *ASTM B627, Std. Test Method for Electrolytic Corrosion Testing (EC Test)*, American Society for Testing and Materials, Philadelphia, Pa., 1992.

45. *ASTM B533, Std. Test Method for Peel Strength of Metal Electroplated Plastics*, American Society for Testing and Materials, Philadelphia, Pa., 1992.

46. Ref. 19, p. 49.

47. J. Dini, *ASTM STP947, Ring Shear Testing of Deposited Coatings*, American Society for Testing and Materials, Philadelphia, Pa., 1987.

48. Ref. 19, pp. 254–261.

49. *ASTM B571, Std. Test Methods for Adhesion of Metallic Coatings*, American Society for Testing and Materials, Philadelphia, Pa., 1991.

50. *ASTM B489, Std. Practice for Bend Test for Ductility of Electrodeposited and Autocatalytically Deposited Metal Coatings on Metals*, American Society for Testing and Materials, Philadelphia, Pa., 1992.

51. *ASTM B490, Std. Practice for Micrometer Bend Test for Ductility of Electrodeposits*, American Society for Testing and Materials, Philadelphia, Pa., 1992.

52. I. Kim and R. Weil, *Tension Testing of Very Thin Electrodeposits*, in Ref. 47.

53. S. Nakahara, Y. Okinaka, and H. Strashil, *Ductility of Plated Metal Films: Its Measurement and Relationship to Microstructure*, in Ref. 47.

54. *ASTM B578, Std. Test Method for Microhardness of Electrodeposited Coatings*, American Society for Testing and Materials, Philadelphia, Pa., 1987.

55. J. Homer, *Microhardness Testing of Plating Coatings: Recent Round-Robin Experiences*, in Ref. 47.

56. J. Homer, "Microhardness Testing of Plating Coatings: Defining Precision and Bias," *Int'l Tech Conf. Proc.*, AESF SUR FIN, Atlanta, Ga., 1992.

57. *ASTM B735, Std. Test Method for Porosity in Gold Coatings on Metal Substrates by Nitric Acid Vapor*, American Society for Testing and Materials, Philadelphia, Pa., 1992.

58. *ASTM B741, Std. Test Method for Porosity in Gold Coatings on Metal Substrates by Paper Electrogaphy*, American Society for Testing and Materials, Philadelphia, Pa., 1990.

59. *ASTM B809, Std. Test Method for Porosity in Metallic Coatings by Humid Sulfur Vapor (flowers-of-sulfur)*, American Society for Testing and Materials, Philadelphia, Pa., 1990.

60. *ASTM B 765, Std. Guide for Selection of Porosity Tests for Electrodeposits and Related Coatings*, American Society for Testing and Materials, Philadelphia, Pa., 1993.

61. Ref. 26, p. 146.

62. W. Safranek, *Properties of Electrodeposited Metals and Alloys*, 2nd ed., AESF Publication, Orlando, Fla., 1986.

63. *ASTM B733, Std. Spec. for Autocatalytic Nickel–Phosphorus Coatings on Metals*, American Society for Testing and Materials, Philadelphia, Pa., 1990.

64. D. Berman, "Effect of Baking and Stress on the Hydrogen Content of Cadmium Plated High Strength Steels," paper 192 presented at *Corrosion/85 NACE, Materials Performance*, Boston, Mass., Sept. 1985.

65. Ref. 19, pp. 176–196.

66. *ASTM F326, Std. Test Method for Electronic Hydrogen Embrittlement Test for Cadmium Electroplating Processes*, American Society for Testing and Materials, Philadelphia, Pa.

67. *ASTM F519, Std. Method for Mechanical Hydrogen Embrittlement Testing of Plating Processes and Aircraft Maintenance Chemicals*, American Society for Testing and Materials, Philadelphia, Pa., 1977.

68. *ASTM STP 543, Hydrogen Embrittlement Testing*, American Society for Testing and Materials, Philadelphia, Pa.

69. Ref. 19, pp. 199–219.

70. A. Brenner and S. Senderoff, *J. Res. Nat. Bur. Stds.* **42**(2) (1949).

71. *ASTM B636, Std. Method for Measurement of Internal Stress of Plated Metallic Coatings with the Spiral Contractometer*, American Society for Testing and Materials, Philadelphia, Pa., 1992.

72. Ref. 19, pp. 214–216.

73. K. Parker, *Internal Stress Measurements of Electroless Nickel Coatings by the Rigid Strip Method*, ASTM STP 947, American Society for Testing and Materials, Philadelphia, Pa., 1987.

74. F. Altmayer, *Pltg. Surf. Fin.*, 54–56, (June 1992).

75. Adapted from Ref. 5, p. 777.

76. S. Nutt and S. Zaidi, *Pltg. Surf. Fin.*, (Apr. 1985).

77. Ref. 5, p. 806.

78. Ref. 5, pp. 520–551.

79. T. Irvine, *Chemical Analysis of Electroplating Solutions*, Chemical Publishing Co., New York, 1970.

80. P. Wild, *Modern Analysis for Electroplating*, 2nd ed., Finishing Publications, Ltd., Middlesex, UK, 1984.

81. H. Geduld, *Zinc Plating*, ASM Int'l. and Finishing Publications Ltd., Materials Park, Ohio, 1988, pp. 265–293.

82. W. Nohse, *The Hull Cell*, Robert Draper Ltd., Teddington, UK, 1966.

83. Ref. 7, pp. 30–31.

84. Ref. 26, p. 442.

85. *ASTM B766, Std. Spec. for Electrodeposited Coatings of Cadmium*, American Society for Testing and Materials, Philadelphia, Pa., 1986.

86. *ISO 2082, Electroplating Coatings of Cadmium on Iron and Steel*, ISO, Geneva, Switzerland, 1986.

87. C. Peger, *Porous Pots*, Chrome Plating Consultants, Inc., Cleveland, Ohio, 1992.

88. Ref. 5, p. 192.

89. *ASTM B456, Std. Spec. for Electrodeposited Coatings of Copper Plus Nickel Plus Chromium and Nickel Plus Chromium*, American Society for Testing and Materials, Philadelphia, Pa., 1993.

90. D. Snyder, *Metal Finishing* **90**, 6 (June 1992).

91. Ref. 5, p. 179.

92. **U.S. Pat. 1,985,308** (Dec. 25, 1934). O. Bornhauser.

93. *ASTM B177, Std. Practice for Chromium Electroplating on Steel for Engineering Use*, American Society for Testing and Materials, Philadelphia, Pa., 1993.

94. *ASTM B650, Std. Spec. for Electrodeposited Engineering Chromium Coatings of Ferrous Substrates*, American Society for Testing and Materials, Philadelphia, Pa., 1985.

95. Ref. 17, p. 144.

96. *Cobalt Sulfamate 20*, operating data sheet, Allied-Kelite Div., Witco, Melrose Park, Ill., 1980.

97. J. Homer, "A Study of the Effects of Some Variables on the Speed and Distribution of Deposits from Cyanide Copper Plating Solutions," *51st Annual Tech Proceedings, St. Louis, Mo.*, American Electroplaters' Society, Orlando, Fla., 1964.

98. Ref. 5, p. 202.

99. Ref. 26, p. 314.

100. *ASTM B604, Std. Spec. for Decorative Electroplated Coatings of Copper/ Nickel/ Chromium on Plastics*, American Society for Testing and Materials, Philadelphia, Pa., 1991.

101. A. Brenner, *Electrodeposition of Alloys*, Vol. 1, Academic Press, Inc., New York, 1963, pp. 457–469.

102. "Brass Plating," in Ref. 101, chapter with over 200 references.

103. "Brass Plating," in Ref. 17, chapter with 190 references.

104. "Bronze Plating," in Ref. 101, chapter with 90 references.

105. *ASTM B488, Std. Spec. for Electrodeposited Coatings of Gold for Engineering Uses*, American Society for Testing and Materials, Philadelphia, Pa., 1986.

106. A Weisberg, in Ref. 5.

107. Indium Corp. of America, operating data sheets, 1992.

108. **U.S. Pat. 2,318,592** (May 11, 1943), M. Cupery (to E.I. du Pont de Nemours and Co., Inc.).

109. R. Piontelli, *J. Electrochem. Soc.* **94**, 106 (1948).

110. *ASTM B200, Std. Spec. for Electrodeposited Coatings of Lead and Lead-Tin Alloys on Steel and Ferrous Alloys*, American Society for Testing and Materials, Philadelphia, Pa., 1985.

111. O. Watts, *Trans. Amer. Electrochem. Soc.* **29**, 395 (1916).

112. R. Piontelli and Cambi, *Rend. Inst. Lombard Sci.* **72**, 128 (1938).

113. R. Barrett, *Electrotypers and Stereotypers Bulletin* **36**, 33 (1950).

114. **U.S. Pat. 4,436,591** (Mar. 13, 1989), to J. A. de Hek (to Veco Beheer, Netherlands).

115. *Bulletins on Ni-Speed Process*, INCO, Toronto, Canada, 1992.

116. *ASTM B503, Std. Practice for Use of Copper and Nickel Electroplating Solutions for Electroforming*, American Society for Testing and Materials, Philadelphia, Pa., 1984; Ref. 57, Chaps. 13, 14, and 15.

117. **U.S. Pat. 1,818,229** (Aug. 11, 1931), G. Lutz and L. Westbrook (to Grasselli Chemical Co.).

118. *Proceedings of Conference on Theory and Practice of Bright Electroplating*, Vilnius, Lithuania Dec. 1962, trans. by U.S. Dept. of Commerce Clearinghouse, Washington, D.C.

119. *ASTM B764, Std. Method for Simultaneous Thickness and Electrochemical Potential Determination of Individual Layers in Multi-layer Nickel Deposit (STEP test)*, American Society for Testing and Materials, Philadelphia, Pa., 1986.

120. **U.S. Pat. 2,844,530** (July 22, 1958), W. Wesley and B. Knapp (to INCO).

121. A Brenner, *Electrodeposition of Alloys*, Vols. 1-3, Academic Press, Inc., New York, 1963.

122. **U.S. Pat. 2,026,718** (Jan. 7, 1936), L. Weisberg and W. Stoddard (to Weisberg and Greenwald, Inc.).

123. Ref. 114, Vol. 1, p. 261.

124. A. Brenner, D. Couch, and E. Williams, *J. Research Nat'l Bur. Standards* **44**, 109-122 (1950).

125. Ref. 121, Vol. 1, p. 397.

126. J. Horner, unpublished work, Allied-Kelite Div., Witco, New Hudson, Mich., 1993.

127. **U.S. Pat. 4,892,627** (Jan. 9, 1990), K. Takada (to Takada Laboratories.)

128. *ASTM B634, Std. Spec. for Electrodeposited Coatings of Rhodium for Engineering Use*, American Society for Testing and Materials, Philadelphia, Pa., 1993.

129. R. Morrissey, in Ref. 5.

130. J. Notley, *Trans. Inst. Met. Finishing* **50**, 58-62 (1972).

131. A. Blair, in Ref. 5.

132. *ASTM B700, Std. Spec. for Electrodeposited Coatings of Silver for Engineering Use*, American Society for Testing and Materials, Philadelphia, Pa., 1990.

133. *ASTM B545, Std. Spec. for Electrodeposited Coatings of Tin*, American Society for Testing and Materials, Philadelphia, Pa., 1992.

134. D. Swalheim, in Ref. 17, p. 374.

135. N. Parkinson, *J. Electrodepositors Tech. Soc.* **27**, 129-151 (1951).

136. *ASTM B605, Std. Spec. for Electrodeposited Coatings of Tin-Nickel Alloy*, American Society for Testing and Materials, Philadelphia, Pa., 1982.

137. *ISO 2179, Electrodeposited Coatings of Tin-Nickel Alloy*, Specification and Test Methods, ISO, Geneva, Switzerland, 1986.

138. T. Natorski, *Metal Finishing* **90**(3), 15 (Mar. 1992).

139. Unpublished results of market survey, Allied-Kelite Div., Witco, Melrose Park, Ill., Feb. 1992.

140. D. Crotty, *Metal Finishing* **89**(11), 58 (1991).

141. S. Hoyt, *Metal Data*, Reinhold Publishing Co., New York, 1952, p. 450.

142. J. Horner, unpublished work, Allied Research Laboratories, Detroit, Mich., 1992.

143. *ASTM B633, Std. Spec. for Electrodeposited Coatings of Zinc on Iron and Steel*, American Society for Testing and Materials, Philadelphia, Pa., 1985.

144. *ASTM B487, Std. Test Method for Measurement of Metal and Oxide Coating Thickness by Microscopical Examination of a Cross Section*, American Society for Testing and Materials, Philadelphia, Pa., 1990.

145. *ASTM B499, Std. Test Method for Measurement of Coating Thicknesses by the Magnetic Method: Nonmagnetic Coatings on Magnetic Basis Metals*, American Society for Testing and Materials, Philadelphia, Pa., 1988.

146. *ASTM B253, Std. Guide for Preparation of Aluminum Alloys for Electroplating*, American Society for Testing and Materials, Philadelphia, Pa., 1987.

147. *ASTM B480, Std. Guide for Preparation of Magnesium and Magnesium Alloys for Electroplating*, American Society for Testing and Materials, Philadelphia, Pa., 1988.

148. **U.S. Pat. 2,729,601** (Jan. 3, 1956), J. Beach and C. Faust (to United States of America, as represented by the Atomic Energy Commission).

149. *ASTM B695, Std. Spec. for Coatings of Zinc Mechanically Deposited on Iron and Steel*, American Society for Testing and Materials, Philadelphia, Pa., 1991.

150. G. Hsu, *Pltg. Surf. Fin.*, 52-55, (Apr. 1984).

151. S. Watson, *Trans. Inst. Met. Finishing* **70**(1), 28 (1992).

152. *Nickel Currents, An International Newsletter*, Vol. 1, No. 1, INCO, Toronto, Canada, Mar. 1993.

153. C. Smith, "Investigation of Nickel-Cadmium-Diffused and Cadmium Plating Replacement Coatings," paper 920943 presented at *SAE Airlines Plating Forum*, San Diego, Calif., 1992.

154. E. Gomez and co-workers, *Metal Finishing*, 87-91, (June 1992).

155. C. Huang, *Pltg. Surf. Fin.*, 64-67 (Dec. 1989).

156. N. Short, A. Abibsi, and J. Dennis, *Trans. Inst. of Met. Fin.*, 67-73 (Aug. 1989).

157. Ref. 7, pp. 413-425.

158. *MIL C-5541, Chemical Films for Aluminum and Aluminum Alloys*, U.S. Government Printing Office, Washington, D.C., 1990.

159. *MIL M-3171, Processes for Pretreatment and Prevention of Corrosion of Magnesium Alloys*, U.S. Government Printing Office, Washington, D.C., 1974.

160. J. Horner, "Chromate Coatings for Zinc," *AES Zinc Symposium*, Oct. 1977, AESF, Orlando, Fla.

161. Technical data sheets, *Zinc alloy processes*, Enthone-OMI, Inc., New Haven, Conn., 1992.

162. *ASTM B431, Std. Practice for Processing Mandrels for Electroforming*, American Society for Testing and Materials, Philadelphia, Pa., 1985.

Jack Horner
Allied-Kelite

ELECTROSEPARATIONS

Electrodialysis,
Electrophoresis,

ELECTRODIALYSIS

Electrodialysis (ED) is a process for moving ions across a membrane from one solution to another under the influence of a direct electric current (see DIALYSIS). Classically the process was carried out in three-compartment electrolytic cells in which the compartments were separated from each other by essentially nonselective membranes (see MEMBRANE TECHNOLOGY). The end compartments contained electrodes. In 1940, a multicompartment ED process using ion-selective membranes (Fig. 1) was suggested (1) in which membranes *A* selective to anions alternated with membranes *C* selective to cations. When a d-c potential is applied, cations M^+ tend to move toward the negatively charged cathode. These ions are able to permeate the cation-selective membranes but not the anion-selective membranes if the latter are perfectly selective. Similarly, anions X^- tend to move toward the positively charged anode. These ions are able to permeate the anion-selective membranes but not the cation-selective ones. As a result, the odd-numbered compartments in the figure become depleted in electrolyte, the even-numbered compartments enriched.

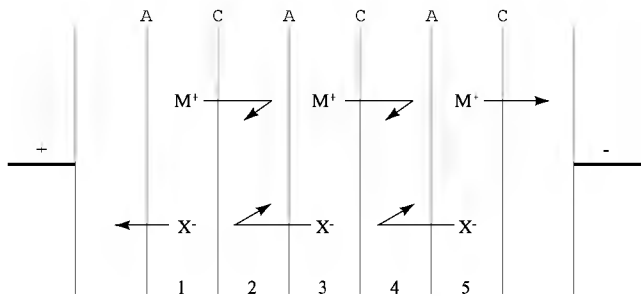


Fig. 1. Principle of multicompartiment electrodialysis. See text.

Membranes

In 1950, ion-selective membranes having high selectivity, low electrical resistance, good mechanical strength, and good chemical stability were described (2). These were essentially insoluble, synthetic, polymeric, organic ion-exchange (IX) resins in sheet form (see ION EXCHANGE). Typical chemical structures for modern membranes of this type are shown schematically in Figure 2. The cation-selective membranes (Fig. 2a) consist of cation-exchange (CX) resin composed of polystyrene having negatively charged sulfonate groups chemically bonded to most of its phenyl groups. The charges of the sulfonate groups are electrically balanced by positively charged cations (counterions). Sulfonated polystyrene swells greatly in water. The amount of swelling is typically controlled by including cross-linking agents in the polymer, represented by divinylbenzene in Figure 2; by incorporating electrically neutral polymers; or by having extensive regions (blocks) in the polymer which lead to substantial microcrystallinity. The positively charged counterions, eg, Na^+ , Ca^{2+} , or Mg^{2+} , are appreciably dissociated from the chemically bound negatively charged groups once the membrane is exposed to water. Thus, in water, the counterions are mobile, and may be exchanged for other cations from an ambient solution, maintaining the electrical neutrality of the membrane. This high (typically $> 1 \text{ meq cm}^{-3}$ of membrane) concentration of counterions in IX resins is responsible for the low electrical resistance of the membrane. The high concentration of bound negatively charged groups tends to exclude mobile negatively charged ions (co-ions) from an ambient solution and is responsible for the high ion selectivity of the membranes.

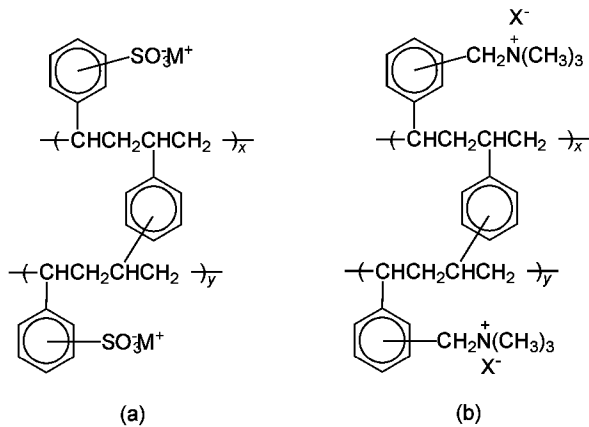


Fig. 2. Schematic representation of (a) cation-exchange resin, and (b), anion-exchange resin.

The anion-selective (AX) membranes (Fig. 2b) also consist of cross-linked polystyrene but have positively charged quaternary ammonium groups chemically bonded to most of the phenyl groups in the polystyrene instead of the negatively charged sulfonates. In this case the counterions are negatively charged, eg, Cl^- , HCO_3^- , NO_3^- , or SO_4^{2-} .

Commercially available membranes are usually reinforced with woven, synthetic fabrics to improve the mechanical properties. Several hundred thousand square meters of IX membranes are now produced annually, and the mechanical and electrochemical properties are varied by the manufacturers to suit the proposed applications. The electrochemical properties of most importance for ED are (1) the electrical resistance per unit area of membrane; (2) the ion transport number, related to current efficiency; (3) the electrical water transport, related to process efficiency; and (4) the back-diffusion, also related to process efficiency.

Commercial IX membranes have thicknesses of ca 0.15–0.5 mm and electrical resistances of ca 3–20 $\Omega\text{-cm}^2$ at 25°C when in equilibrium with 0.5 N sodium chloride. The electrical resistances are somewhat higher in more dilute solutions because co-ions are more effectively excluded from the membrane by the IX resin. The electrical resistance decreases with increasing temperature at a rate of ca -1.9% /°C. The electrical resistance of an ED apparatus

depends in large part on the electrical resistances of the membranes when electrolyte solutions are in excess of about 0.1 N. For more dilute solutions the resistance of the apparatus tends to be dominated by the resistance of the solution being demineralized.

The ion transport number is defined as the fraction of current carried through the membrane by counterions. If the concentration of fixed charges in the membrane is high compared to the concentration of the ambient solution, then the mobile ions in the IX membrane are mostly counterions, co-ions are effectively excluded, and the ion transport number then approaches 1. Commercial membranes have ion transport numbers in dilute solutions of ca 0.85–0.95. The relationship between ion transport number and current efficiency is shown in Figure 3 where $\bar{t}^A$ is the fraction of current carried by the counterions (anions) through the AX membrane and $\bar{t}_+^C$ is the fraction of current carried by the counterions (cations) through the CX membrane. The remainder of the current $(1 - \bar{t}^A) = \bar{t}_+^A$ in the case of the AX membranes and $(1 - \bar{t}_+^C) = \bar{t}_-^C$ in the case of the CX membranes is carried by co-ions and constitutes an electrical inefficiency. The net transport of electrolyte is then given by:

$$\bar{t}^A - (1 - \bar{t}_+^C) = \bar{t}^A - \bar{t}_+^C = \bar{t}_-^A - \bar{t}_+^C$$

The electrical water transport is defined as water that accompanies the electrical transport of ions through the membranes. Owing to the high concentration of counterions in IX membranes, the transport of ions and water are closely coupled, ie, the internal solution tends to move in piston-like flow. The electrical water transport is ca 100–200 cm³ eq of ions transferred. This value is toward the low end of the range for AX membranes and toward the high end for CX membranes. In the case of dilute solutions, water transport is not of significant engineering importance. However, in more concentrated solutions, such as seawater, the water transport can be a significant fraction of the volume of solution electrodyalized and, therefore, constitutes a process inefficiency. Generally, membranes having low water contents and high concentrations of fixed, charged groups have low electrical water transport. Such membranes also tend to have high ion transport numbers and are therefore preferred for concentrated electrolytes.

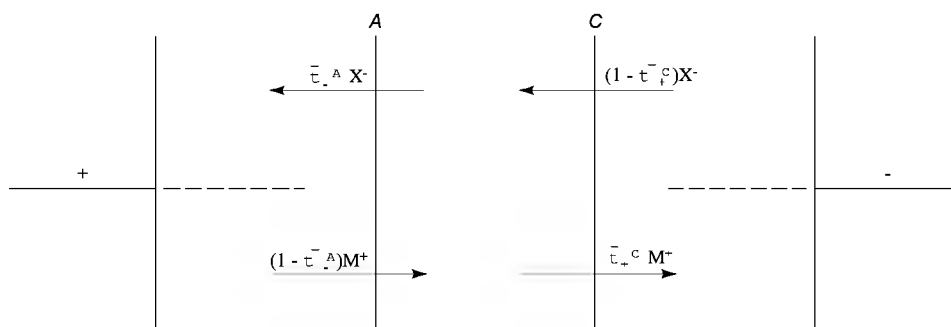


Fig. 3. Relationship between current efficiency and ion transport number.

Back-diffusion is the transport of co-ions, and an equivalent number of counterions, under the influence of the concentration gradients developed between enriched and depleted compartments during ED. Such back-diffusion counteracts the electrical transport of ions and hence causes a decrease in process efficiency. Back-diffusion depends on the concentration difference across the membrane and the selectivity of the membrane; the greater the concentration difference and the lower the selectivity, the greater the back-diffusion. Designers of ED apparatus, therefore, try to minimize concentration differences across membranes and utilize highly selective membranes. Back-diffusion between sodium chloride solutions of zero and one normal is generally $<ca\ 2 \times 10^{-6}$ meq (s·cm²).

Bipolar ED. Bipolar ED membranes have one surface consisting of CX resin and the opposite surface of AX resin. When a direct current is passed through such a membrane in a direction to pull anions out of the interface between the AX and CX resins and through the AX resin, the interface rapidly becomes depleted of all ions other than those resulting from the dissociation of water. Dilute alkali can therefore be produced at the outer surface of the AX region and dilute acid at the outer surface of the CX layer (3). An enormous effort has been applied by the AquaTech division of Allied-Signal Corp. to refine and commercialize this technology, which as of this writing is without signal commercial success.

The bipolar membranes are used in a more or less conventional ED stack together with conventional unipolar membranes. Such a stack has many acid-alkali producing membranes between a single pair of end electrodes. The advantages of the process compared to direct electrolysis seem to be that because only end electrodes are required, the cost of the electrodes used in direct electrolysis is avoided, and the energy consumption at such electrodes is also avoided.

The disadvantages appear to be that the bipolar membranes are comparatively expensive, and the economic life is limited to about one year. Such short lifetime appears to result from the very high ($\sim 10^6$ V·cm) voltage gradients at the interface between the AX and CX regions. Additionally, practical current densities are limited to about 1000 A·m² available area.

Apparatus

The apparatus for ED is fundamentally an array of alternating AX and CX membranes terminated by electrodes. The membranes are separated from each other by gaskets which form fluid compartments. Compartments that have AX membranes on the side facing the positively charged anode are electrolyte-depletion compartments. These are also called demineralizing, diluting, diluate, or dilute compartments. The remaining compartments are electrolyte-enrichment compartments, also called concentrating, concentrate, or brine compartments. The enrichment and depletion compartments also alternate through the array. Holes in the gaskets and membranes register with each other to provide two pairs of internal hydraulic manifolds to carry fluid into and out of the compartments. One pair communicates with the depletion compartments, and the other with the enrichment compartments. Much effort has been expended on the design of the entrance and exit channels from the manifolds to the compartments to prevent unwanted cross-leak of fluid intended for one class of compartment into the other class. This effort has been made increasingly difficult by the trend to thinner membranes and gaskets, the latter determining membrane spacing and the thickness of the fluid compartments; such trends are intended to reduce energy consumption. A contiguous group of two membranes and the associated two fluid compartments is called a cell pair. A group of cell pairs and the associated end electrodes is called a stack or a pack. Generally 100–600 cell pairs are arranged in a single stack, the choice being made on the basis of ED capacity desired, the uniformity of flow distribution achieved among the several compartments of the same class in a stack, and the maximum total direct current potential desired. One or more stacks may be arranged in a press, designed to compress the membranes and gaskets against the force of fluid flowing through the compartments thereby preventing fluid leaks to the outside and internal cross-leaks between compartments. For small presses such compression is usually provided by tierods; for larger presses hydraulic rams are frequently used.

Commercial membranes have typical thicknesses of ca 0.15–0.5 mm; the compartments between the membranes have typical thicknesses of ca 0.5–2 mm. The thickness of a cell pair is therefore in the 1.3–5.0 mm range, commonly about 3.0 mm. One hundred cell pairs have a combined thickness of about 300 mm. The effective area of a cell pair for current conduction is generally on the order of 0.2–2 m².

In concentrated electrolytes the electric current applied to a stack is limited by economic considerations, the higher the current I the greater the power consumption W in accordance with the equation $W = I^2 R_s$, where R_s is the electrical resistance of the stack. In relatively dilute electrolytes the electric current that can be applied is limited by the ability of ions to diffuse to the membranes. This is illustrated in Figure 4 for the case of an AX membrane. When a direct current is passed, a fraction ($t^A \simeq 0.85$ –0.95) is carried by anions passing out of the membrane–solution interface region and through the membrane. In the bulk solution, a fraction t of the current is carried by anions passing into the interfacial region. Generally t is significantly

less than $\bar{t}^A$. For example, in the case of sodium chloride, t^- is ca 0.6. As the electric current continues to pass, the interfacial region becomes depleted in electrolyte. The difference between the quantity of anions transferred out of the interfacial region and those transferred in must be made up by convection and diffusion. If the interfacial region is defined as the region of streamline flow, then diffusion is the only mechanism, other than conduction, available to bring anions into the region. At steady state the concentration of electrolyte in the solution at the membrane surface must be reduced sufficiently from the bulk value to provide the concentration gradient necessary to bring in by diffusion the difference between the anions carried out of and into the interfacial regions by the electric current. This may be expressed by:

$$i\bar{t}^A - it^- = FD(c_b - c) / \delta$$

where F is Faraday's constant 96,500 C, the quantity of electric current required to transfer one equivalent of ions through a perfectly selective membrane; i is the current density in A/cm²; D is the diffusion constant for the electrolyte, cm²/s; c_b is the concentration of the electrolyte in the bulk (nonstreamline region), g-eq/mL; c , also in g-eq/mL, is the concentration of electrolyte in solution at the membrane surface; and δ in cm is the thickness of the interfacial (streamline flow) region. This expression may be rearranged to:

$$i = \frac{DF(c_b - c)}{\delta(\bar{t}^A - t^-)}$$

Hence in any given situation the maximum current density that can be carried by a combination of electrical conduction and diffusion occurs when c , the concentration in the solution at the membrane surface, approaches zero, ie,

$$i_{\max} = \frac{FDc_b}{\delta(\bar{t}^A - t^-)}$$

For a typical electrodialysis apparatus, δ appears to be of the order of 5×10^{-3} cm, and $\bar{t}^A$ is about 0.9. For dilute sodium chloride solutions, D is ca 1.5×10^{-5} cm²/s and t^- about 0.6. Under these circumstances the ratio $i_{\max}/c_b$ is ca 1000 A/cm (g-eq), ie, $i_{\max}$ is ca 0.1 A/cm² when c_b is ca 0.1 N. Increasing temperature increases D by ca 2.3% /°C, and decreasing δ increases $i_{\max}$. Therefore, designers usually include some kind of structure in the ED compartments to break up the streamline interfacial region and bring electrolyte as close as possible to the membrane surface by convection. The maximum current also increases as the difference $(\bar{t}^A - t^-)$ decreases. As $\bar{t}^A$ approaches t^- , the diffusion-controlled limitation on current density vanishes and the current density may be increased without limit. However, power consumption increases as i^2R and puts an economic limitation on current density. When $\bar{t}^A$ approaches t^- , the anion membrane loses its selectivity and becomes essentially neutral. The net transport of electrolyte is then given by $(t^- - \bar{t}^C)$. Thus for many electrolytes, an ED apparatus utilizing alternating neutral and CX membranes exhibits substantial depletion and enrichment (4). For example, in the case of sodium chloride solutions the net transport of electrolyte is 0.8 (current efficiency, 80%) when the ion transport numbers of the AX and CX membranes are each 0.9. If the AX membrane is replaced by a neutral membrane, the net transport is 0.5 (current efficiency, 50%) when the concentration difference between enriched and depleted compartments is small. ED apparatus using neutral membranes has not been commercially successful.

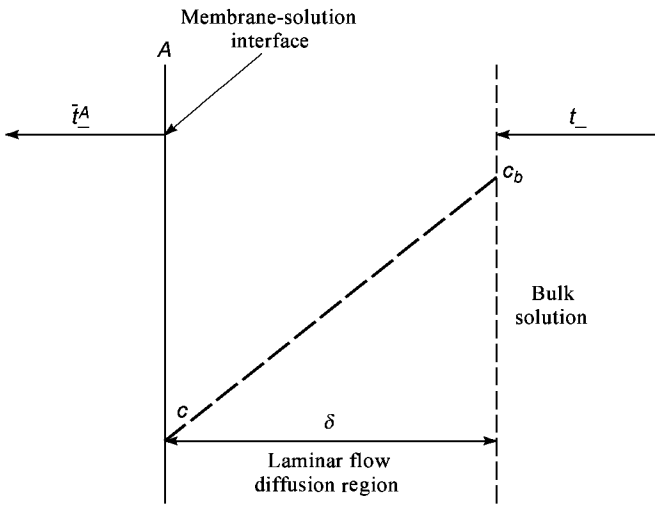


Fig. 4. Current limitation controlled by diffusion to anion membrane.

Polarization. When the applied current density equals $i_{\max}$, the AX membrane and the apparatus are said to be concentration polarized or simply polarized. At $i_{\max}$ the fluid at the surface of the membrane is essentially depleted of electrolyte and the electrical resistance of the apparatus increases substantially even though the bulk solution in the depletion compartments may still contain an appreciable concentration of electrolyte. This is a sign of polarization.

At the polarization current density, ions resulting from the dissociation of water have concentrations comparable to the concentration of electrolyte at the surface of the membrane. A significant fraction of the current through the AX membrane is then carried by hydroxide ions into the enrichment compartments. Hydrogen ions are carried into the bulk solution in the depletion compartments. Changes in the pH of the enrichment and depletion compartments are another sign of polarization.

If $i_{\max}$ is substantially exceeded, electrolytes that are insoluble at high pHs, such as calcium carbonate and magnesium hydroxide, may precipitate at the interface between the AX membrane and the enriched solution. This is a third sign of polarization.

In sodium chloride solutions the ion transport number for Na⁺ is about 0.4 compared to about 0.6 for Cl⁻. Thus a CX membrane would be expected to polarize at lower current densities than an AX membrane. Careful measurements show that CX membranes do polarize at lower current densities; however, the effects on pH are not as significant as those found when AX membranes polarize. Such differences in behavior have been satisfactorily explained as resulting from catalysis of water dissociation by weakly basic groups in the AX membrane surfaces and/or by weakly acidic organic compounds absorbed on such surfaces (5).

Many fluids of natural origin contain detectable quantities of high molecular weight organic anions, such as those from humic, fulvic, and tannic acids, which can be carried to and deposited on AX membranes. Such deposits can behave as thin films partially selective to cations (6). The interfaces between such films and the underlying AX membranes then act as very thin stagnant depletion compartments and the AX membranes may exhibit polarization at current densities that are much lower than would be expected for new membranes in the absence of such anions.

Performance. The performance of an ED stack may be estimated by considering the material balance around the stack:

$$\Delta(vc) = v_i c_i - v_o c_o = \frac{iEA_p N_p}{F}$$

where v_i and v_o are the flow rates in mL s into and out of the depletion compartments of the stack, respectively; c_i and c_o are the concentrations in g-eq mL into and out of the depletion compartments, respectively; $\bar{i}$ is the average current density in A cm²; E is the net ion transfer ($\bar{i}^A - \bar{i}^C$); A_p is the current-carrying area per cell pair, cm²; N_p is the number of cell pairs in the stack; and F is Faraday's constant. Typically a stack is designed so that $v_o c_o$ is ca 50% of $v_i c_i$. This is because $\bar{i}_{\max}$ is determined by c_o . When c_o is very small compared to c_p , then $\bar{i}$ may be uneconomically low. As a matter of conservative engineering practice, the stack is designed so that $\bar{i}$ is appreciably lower than $\bar{i}_{\max}$.

The power consumption W may be estimated from $\bar{i}^2 R_p A_p N_p$ where R_p the electrical resistance of a cell pair per unit area ($\Omega \cdot \text{cm}^2$), is given by:

$$R_p = R_d + R_e + R_C + R_A + R_{df} + R_T$$

R_d is the average resistance of a depletion compartment; R_e is the average resistance of an enrichment compartment; R_C is the average resistance of the CX membranes; R_A is the average resistance of the AX membranes; R_{df} is the resistance contributed by depletion in the membrane-solution interfaces; and R_T is the resistance offered by the concentration potential between the enrichment and depletion compartments, ie, a measure of the thermodynamic work required to achieve the concentration difference.

Generally these individual contributions to R_p are not separately measured or calculated. Instead R_p itself is correlated with $\bar{i}, c_o$ and c_i . For dilute solutions (7):

$$R_p = b + \frac{a}{c_a}$$

where a and b are empirically determined constants and c_a is the average concentration defined as:

$$c_a = \frac{2c_{le}c_{ld}}{c_{le} + c_{ld}}$$

The subscripts have the following significance: ld indicates the log-mean concentration in the depletion compartments; le indicates the log-mean concentration in the enrichment compartments.

Typical values at room temperature are $a, 0.001\text{--}0.0025 \text{ }\Omega \cdot \text{g-eq cm}$; $b, \text{ca } 20\text{--}40 \text{ }\Omega \cdot \text{cm}^2$. The values of a and b within the ranges depend on electrolyte composition and on design of the electrodialysis apparatus. As stated above, R_p decreases with increasing temperature at ca 1.9% °C. Typical d-c voltages are ca 0.5–1 volt per cell pair. Assuming an overall current efficiency of 90% this implies a d-c electrical energy requirement of ca 15–30 W·h g-eq of electrolyte transferred. To this must be added the energy consumption of auxiliary equipment such as pumps and instrumentation, about 500 W·h m³, and energy losses during conversion of ac to dc. In the case of relatively concentrated electrolytes, for economic reasons, designers tend to choose d-c voltages per cell pair from the low end of the range mentioned above.

Uses

High Purity Water. Electrodialysis is being increasingly used in conjunction with other processes to produce pure and ultrapure water for use in high pressure boilers and in electronics and pharmaceutical manufacture (8,9). A typical application minimally involves six steps: (1) potable or brackish water is treated via cross-flow ultrafiltration (UF) (see ULTRAFILTRATION) to remove particulate matter; (2) the water is treated by reversing-type electrodialysis (EDR) to remove about 90% of the electrolytes; (3) the product of electrodialysis is then subjected to reverse osmosis (qv) to remove a substantial fraction of silica and organics, and about 90% of the remaining electrolytes; (4) the permeate from reverse osmosis is further demineralized by electrodeionization (EDI) to remove a further 90 to 95% of the electrolytes still remaining; (5) the water is finally treated by mixed-bed ion-exchange deionization followed by (6) microfiltration. Activated carbon may be included to remove organic materials. Chlorine, chlorine dioxide, ozone, hydrogen peroxide, and or uv irradiation may be used to sterilize the water and or partially oxidize the organics to enable the latter to be more easily removed.

In any given hybrid process train designed for the production of ultrapure water, any one or more of the process steps may be omitted. These trains have very low chemical and operator requirements. By utilizing EDR, which has almost no requirement for chemical additives or chemical cleaning, before reverse osmosis, the requirement of the latter for chemical additives is essentially eliminated. Usage of EDI reduces the need for chemical regeneration of mixed-bed ion-exchange deionizers by about a factor of 10 to 20. These deionizers are nevertheless needed to produce ultrapure water of the highest possible electrical resistivity.

Brackish Water. Production of potable water from brackish water has been one of the principal applications of ED (see WATER SUPPLY AND DESALINATION). The term brackish water is used to define a water that is more saline than potable water but appreciably less so than seawater. Brackish waters readily available worldwide have concentrations of electrolyte of ca 20–200 meq L. Potable water generally has <10 meq L and seawater ca 600 meq L. A typical application would be demineralization from 0.05 to 0.005 N for which, at one volt per cell pair, d-c energy consumption would be expected to be ca 1.35 kWh m³ of potable water. Auxiliaries would add about 0.50 kW·h m³. The specific production rates are ca 0.5 m³ m²·d of effective cell pair area. In 1993 the installed cost was about \$135 per m² for large plants, including the immediate electrodialysis auxiliaries such as a-c d-c conversion equipment, pumps, and instrumentation, but not including the costs of land, buildings, brackish water collection and pretreatment, and potable water storage and distribution. Pretreatment generally consists of filtration to remove suspended particulate matter. Small plants may cost from two to four times as much per square meter.

Direct current enters an ED stack at one end through a positively charged anode, generally a refractory metal such as titanium, coated with a thin film of a noble metal or noble metal oxide (see METAL ANODES). The transition from an electron current in the anode to an ionic current in electrolyte bathing the anode produces hydrogen ions and oxygen from the salts in the brackish water. Because hundreds of cell pairs are usually used in an ED stack, the number of equivalents of hydrogen ions and oxygen is generally <1% of the number of equivalents of electrolyte removed by the membranes. The anodes are usually flushed with some of the brackish water to carry away the anode products. The electric current leaves the stack at the other end through a negatively charged cathode generally also titanium coated with noble metal or noble metal oxide. At the cathode, the electric current produces hydrogen gas and hydroxide ions, again in very small quantities compared to the amount of electrolyte removed through the membranes. However, the hydroxide ions can cause precipitation of calcium carbonate or magnesium hydroxide, or both, resulting from the Ca²⁺ or Mg²⁺ ions in the brackish water.

In some ED plants, strong mineral acids, such as sulfuric or hydrochloric, are added to the brackish water which flushes the cathodes. These acids prevent precipitation but are an additional cost and an inconvenience particularly for small plants.

By using reversing-type or EDR stacks, in which each electrode is alternately anodic and cathodic (10), acid need not be added to the cathode. After passing current through the stack for a period of from about 15 minutes to 24 hours the direction of electric current is reversed for a similar period; that is the electrode that was cathodic in the first half of the cycle and precipitated some alkali-insoluble salts, becomes anodic in the second half. Hydrogen ions generated during the anodic half of the cycle dissolve and separate insoluble salts from the electrode. In the EDR stack, compartments that were enriching during one half of the cycle are depleting during the other half. Appropriately placed automatic valves interchange the entering and exiting fluid streams at the beginning of each half cycle. Several thousand EDR plants have been successfully operated for many years. The periodic reversal also removes foreign substances from the EDR stack per se. Such materials may include alkali-insoluble salts such as calcium carbonate produced by marginal concentration polarization at the interfaces between the depleting fluid and the AX membranes; poorly soluble salts, such as calcium sulfate, which precipitate if the solubility limits are exceeded in the enrichment compartments; and high molecular weight organic anions such as humic and fulvic acids which, if present, may tend to be absorbed on AX membranes.

The success of EDR in water demineralization has apparently resulted from its greater tolerance of particulate and fouling matter compared to reverse osmosis; greater forgivingness of process upsets; greater tolerance for unskilled operators; simplicity in design and construction of EDR stacks compared to reverse osmosis modules; the ability to inspect, clean, or replace one membrane at a time; the existence of a comprehensive global sales and

service network; and a vertically integrated manufacture. The economic life of IX membranes in the demineralization of water is generally in excess of 10 years in well-operated plants.

Trends in ED appear to be reduction in pumping and direct ED energy, increase in electric current density, and the use of EDR and hybrid processes in plants in which the manufacturer of the demineralization plant owns and operates the plant, selling water to the user or water distributor.

Concentration of Seawater by ED. In terms of membrane area, concentration of seawater is the second largest use. Warm seawater is concentrated by ED to 18 to 20% dissolved solids using membranes with monovalent-ion-selective skins. The EDR process is not used. The osmotic pressure difference between about 19% NaCl solution and partially depleted seawater is about 20,000 kPa (200 atm) at 25°C, which is well beyond the range of reverse osmosis. Salt is produced from the brine by evaporation and crystallization at seven plants in Japan and one each in South Korea, Taiwan, and Kuwait. A second plant is soon to be built in South Korea. None of the plants are justified on economic grounds compared to imported solar or mined salt.

Industrial Wastes. Closely related to seawater concentration is the simultaneous concentration of industrial effluents and recycle of recovered water (see WASTES, INDUSTRIAL). These applications are expected to increase as environmental restrictions increase. Examples are the concentration of blowdown from cooling towers in power plants; concentration of reverse osmosis blowdown; and the processing of metal treatment wastes (11) (see METAL TREATMENTS).

Electrodeionization. Electrodeionization (EDI) (12) was developed in the 1960s (13) and has been pursued more recently (14). At least the demineralization compartments of ED apparatus are filled with AX and or CX beads or fibers or the IX membranes are embossed and in contact with each other. In the reversing type of EDI, both compartments are so filled. At low current densities the filling acts to augment the surface area of the membranes resulting in a very large increase in allowable current density. At high current densities water splitting occurs, as in the case of normal ED, and the IX filling is converted largely to the hydroxide and hydrogen forms, respectively. The apparatus is then essentially IX regenerated in real time electrochemically.

The EDI process has begun to find commercial applications, almost always as part of a hybrid process for producing demineralized water (8). When essentially 18 MΩ water is required, EDI reduces the load on mixed-bed IX by an order of magnitude or more. The need for regeneration and regeneration chemicals is reduced by about the same factor. Applications are for feed water for high pressure boilers and for ultrapure water for the electronics industry. When less than 18 MΩ water is satisfactory, EDI can entirely replace mixed-bed IX. Typically EDI is preceded in a hybrid plant by some combination of EDR, RO, UF or MF, IX water softening and or absorbent for organics. The tendency of demineralized water producers to reduce chemical requirements seems to assure a future for EDI.

Miscellaneous Applications. The largest miscellaneous application of ED is the demineralization of whey and nonfat milk for food and feed applications (see DAIRY SUBSTITUTES; FEEDS AND FEED ADDITIVES; MILK AND MILK PRODUCTS). During the manufacture of cheese or of milk casein a serum is often produced which contains much of the albumin, globulin, lactose, and minerals present in the milk plus acid and or salt added during manufacture. The albumin and globulin are a valuable food or feed sources but use is generally limited by the electrolytes present. Since the 1960s, ED and EDR have been used to remove excessive electrolytes, generally from concentrated serum. There are approximately 70 such ED or EDR plants worldwide having about 35,000 m² total installed membrane area and producing about 150,000 t yr of demineralized whey solids. Some plants remove in excess of 90% of the ash content of whey. Others remove only roughly 50% of such ash, the remainder being removed by strong acid cation exchange followed by weak base anion exchange. Much of the highly demineralized product is used as a component of mothers' milk replacement.

Other applications include (1) recovery of valuable components from metal plating or treating effluents, including the recovery of hydrofluoric and nitric acids by bipolar ED from stainless steel spent pickle liquor (see ELECTROPLATING; METAL SURFACE TREATMENTS); (2) deashing of beet, cane, or other sugar (qv) juices and molasses. AX membranes are generally subject to fouling by medium molecular weight organic carboxylic acids in such solutions. Although fouling resistant AX membranes have been proposed, the generally short processing season for such sugar solutions leads to high capital charges. Further the additional sugar which could be recovered through desalting is not justified in view of the agricultural policies of the sugar producing countries. There are few, if any, ED plants desalting sugar solutions; (3) deacidification acidification of fruit juices (qv). There are a few plants on a commercial scale; (4) desalting of soy sauce, amino acid solutions, fermentation (qv) products, etc (see AMINO ACIDS). There are a few small ED plants in such applications; and (5) demineralization of blood plasma (see FRACTIONATION, BLOOD PLASMA FRACTIONATION). There are few such plants.

Economic Aspects

The first commercial ED apparatus was sold in 1954 and installed in Saudi Arabia for desalting brackish water. Since then more than 5000 ED plants have been installed worldwide for the demineralization of brackish and potable water. These range in capacity from a few to more than 10,000 m³ d.

Ionics Incorporated (Watertown, Massachusetts), the leading ED supplier, has sold more than 2000 ED and EDR plants having a combined capacity of probably more than 600,000 m³ d. These were originally furnished with probably more than 1.2 × 10⁶ m² of membrane. More than 2000 other ED and EDR plants have been built and installed in China. These plants have a combined capacity of more than 600,000 m³ d and were originally furnished with probably more than 1.2 × 10⁶ m² of membrane. Also it is estimated that more than 1000 ED and EDR plants have been built and installed in the CIS. These plants have a combined capacity of probably more than 300,000 m³ d, probably originally furnished with more than 600,000 m² of membrane. Roughly 100 plants have been installed by other companies such as Corning France (formerly S.R.T.I.) and Portals Water Treatment Ltd. (formerly Permutit-Boby).

NOMENCLATURE

<i>a</i>	empirically determined constant representing that portion of cell pair resistance that is inversely proportional to the average concentration in the cell pair, ohm-cm
<i>A_p</i>	current carrying area per membrane, cm ²
<i>b</i>	empirically determined constant representing that portion of cell pair resistance that is independent of the average concentration in the cell pair, ohm-cm ²
<i>c</i>	concentration, eq mL
<i>c_a</i>	average concentration in the cell pair, eq mL, harmonic mean
<i>c_b</i>	bulk concentration at any point in a depletion cell, eq mL
<i>c_i</i>	concentration of solution inlet to depletion compartment, eq mL
<i>c_d</i>	log-mean concentration in depletion compartment, eq mL
<i>c_e</i>	log-mean concentration in enrichment compartment, eq mL
<i>c_o</i>	concentration of solution from the outlet of a depletion compartment, eq mL
<i>D</i>	diffusion constant of an electrolyte, cm ² s
<i>E</i>	net ion transfer or transport, dimensionless
<i>F</i>	Faraday's constant, about 96,500 C gram-eq
<i>i</i>	current density, A cm ²
$\bar{i}$	average current density, A cm ²
<i>I</i>	current, A
<i>N_p</i>	number of cell pairs in a stack, dimensionless
<i>R_A</i>	average areal resistance of AX membrane, ohm-cm ²
<i>R_C</i>	average areal resistance of CX membrane, ohm-cm ²
<i>R_d</i>	average areal resistance of depletion compartment, ohm-cm ²

R_{df}	average areal resistance contributed by depletion in the membrane-solution interfaces, ohm·cm ²
R_p	average areal resistance of cell pair, ohm·cm ²
R_i	electrical resistance of stack, ohms
R_T	average areal resistance offered by the concentration potential between the enrichment and depletion compartments, ohm·cm ²
t	fraction of current carried by anions in the bulk solution in the depletion compartment, dimensionless
$\bar{t}_+^A$	fraction of current carried through an AX membrane by cations (co-ions), dimensionless
$\bar{t}_-^A$	fraction of current carried through an AX membrane by anions (counterions), dimensionless
$\bar{t}_+^C$	fraction of current carried through CX membrane by cations (counterions), dimensionless
$\bar{t}_-^C$	fraction of current carried through CX membrane by anions (co-ions), dimensionless
v	volumetric flow rate of solution, mL s
v_i	volumetric flow rate of solution entering a compartment, mL s
v_o	volumetric flow rate of solution out of a compartment, mL s
δ	thickness of interfacial, streamline flow region, cm

BIBLIOGRAPHY

"Electrodialysis" in *ECT* 1st ed., Vol. 5, pp. 20–26, by P. Stamberger, Technical Consultant; in *ECT* 2nd ed., Vol. 7, pp. 846–865, by W. K. W. Chen, Celanese Plastics Co.; in *ECT* 3rd ed., Vol. 8, pp. 726–738, by W. A. McRae, Ionics, Inc.

1. K. H. Meyer and W. Strauss, *Helv. Chim. Acta* **23**, 795 (1940).
2. W. Juda and W. A. McRae, *J. Am. Chem. Soc.* **72**, 1044 (1950).
3. V. J. Frlette, *J. Phys. Chem.* **60**, 435 (1956).
4. **U.S. Pat. 2,872,407** (Feb. 3, 1959), P. Kollsman
5. R. Simons, *Electrochimica Acta* **30**, 275 (1985).
6. G. Grossman and A. A. Sonin, *Desalination* **10**, 157 (1972).
7. E. A. Mason and T. A. Kirkham, *Chem. Eng. Prog. Symp. Ser.* **55**, 173 (1959).
8. T. Williamson, G. Coker, K. J. Sims, L. Zhang, and D. Elyanow, *53rd Annual Meeting, International Water Conference*, Pittsburgh, Pa., Oct. 19–21, 1992.
9. H. C. Valcour, Jr., *52nd Annual Meeting, International Water Conference*, Pittsburgh, Pa., Oct. 21–23, 1991.
10. **U.S. Pat. 2,863,813** (Dec. 9, 1958), W. Juda and W. A. McRae.
11. L. R. Schmauss, *1988 Biennial Conference, National Water Supply Improvement Association*, San Diego, Calif., Aug. 1988.
12. W. R. Walters, D. W. Weiser, and L. J. Marek, *Indus. Eng. Chem.* **47**, 61 (1955).
13. **U.S. Pat. 3,149,061** (Sept. 15, 1964), E. J. Parsi.
14. **U.S. Pat. 4,632,745** (Dec. 30, 1986), A. J. Giuffrida, A. D. Jha and G. C. Ganzi.

General References

J. R. Wilson, ed., *Demineralization by Electrodialysis*, Butterworths, London, 1960.

L. H. Shaffer and M. S. Mintz, in K. S. Spiegler, ed., *Principles of Desalination*, Academic Press, Inc., New York, 1966, pp. 199–289.

R. Rautenbach and R. Albrecht, *Membrane Processes*, John Wiley & Sons, Inc., New York, 1989.

K. S. Spiegler, in R. H. Perry and D. Green, eds., *Perry's Chemical Engineers' Handbook*, 6th ed., McGraw-Hill Book Co., New York, 1984, pp. 17–37.

H. Strathmann, *Membrane Separation Systems—A Research Needs Assessment*, U.S. Dept. of Energy, Washington, D.C., p. 8–1.

Wayne A. McRae
Consultant

ELECTROPHORESIS

Electrophoresis is a separation technique most often applied to the analysis of biological or other polymeric samples. It has frequent application to analysis of proteins (qv) and DNA fragment mixtures (see NUCLEIC ACIDS). The high resolution of electrophoresis has made it a key tool in the advancement of biotechnology (qv). Variations of this methodology are being used for DNA sequencing (see GENETIC ENGINEERING), isolating active biological factors associated with diseases such as cystic fibrosis, sickle-cell anemia, myelomas, and leukemia, and establishing immunological reactions between samples on the basis of individual compounds (see also CHEMOTHERAPEUTICS, ANTICANCER; IMMUNOASSAY). Electrophoresis is an extremely effective analytical tool because it does not affect a molecule's structure, and it is highly sensitive to small differences in molecular charge and mass.

The term electrophoresis refers to the movement of a solid particle through a stationary fluid under the influence of an electric field. The study of electrophoresis has included the movement of large molecules, colloids (qv), fibers (qv), clay particles (see CLAYS), latex spheres (see LATEX TECHNOLOGY), basically anything that can be said to be distinct from the fluid in which the substance is suspended. This diversity in particle size makes electrophoresis theory very general.

The fundamental principle behind electrophoresis is the existence of charge separation between the surface of a particle and the fluid immediately surrounding it. An applied electric field acts on the resulting charge density, causing the particle to move, the fluid around the particle to move, or both. An applied electric field also generates heat, through resistive heating, and gases, through electrolysis reactions. Each is important in understanding and designing working electrophoresis equipment.

There are three distinct modes of electrophoresis: zone electrophoresis, isoelectric focusing, and isotachopheresis. These three methods may be used alone or in combination to separate molecules on both an analytical (μL of a mixture separated) and preparative (mL of a mixture separated) scale. Separations in these three modes are based on different physical properties of the molecules in the mixture, making at least three different analyses possible on the same mixture.

Distinction is also made among electrophoretic techniques in terms of the type of matrix employed for analysis. Matrices include polymer gels such as agarose and polyacrylamide, paper, capillaries, and flowing buffers. Each matrix is used for different types of mixtures, and each has unique advantages.

There are a variety of techniques for detecting separated sample compounds using chemical stains, photographic media, and immunochemistry. Each detection technique also gives different information about the identity, quantity, and physical properties of the molecules in the mixture. Detection is often the focus of electrophoresis, and usually yields basic information about the mixture being studied.

Principles

Electrophoresis uses the force of an applied electric field to move molecules or particles, often through a polymer matrix. The electric field acts on the intrinsic charge of a substance, and the force on each substance is proportional to the substance's charge or surface potential. The resulting force on the substance results in a distinct velocity for the substance that is proportional to the substance's surface potential. If two different substances have two different velocities, an electric field applied for a fixed amount of time results in different locations on the matrix for these substances.

The application of an electric field to a gel matrix or capillary tube results in heating in the media and gassing at the electrodes. Thus special attention in the design and use of electrophoretic equipment is required.

Theory of Electrophoretic Motion. The study of the mechanics of electrophoresis focuses on the basis of electric potential on the surface of an object, and the relation of the electric potential to the velocity of the particle. Whereas research has been generally limited to nonmolecular particles of well-defined geometry and is not strictly applicable to molecules such as proteins and DNA fragments, this work is useful for understanding the physics of electrophoretic motion.

The Electric Double Layer. Any time an interface between two immiscible phases occurs, an electric potential can be developed at that interface. For example, the walls of a glass beaker have an electric potential when the beaker holds salt water. The potential is generated because some ions preferentially bind to or absorb onto glass, and the glass naturally has silanol groups, SiO, on its surface that ionize in water causing the glass surface to be charged relative to the bulk salt water. Conceptually, there is a separation of charge that exists in a thin section of the glass and in a thin layer of the salt water adjacent to the glass. These two layers of charge are called the electric double layer. Detailed discussions of the properties of ions and the occurrence of electric double layers may be found in the literature (1–4). In electrokinetic phenomena, the primary concern is with the fluid half of the double layer (in the previous example, the layer is in the salt water), because the layer on the glass itself is immobilized. This charged layer is called the diffuse layer. The electric potential in the diffuse layer extends into the fluid phase, and drops off as a Poisson distribution:

ϕ = ϕ_oexp(- κx)

(1)

where ϕ_o = potential at interface, mV; ϕ = potential × cm interface, mV; κ = characteristic from the length, cm⁻¹; and x = distance from the interface, cm. The characteristic length that the potential extends into the bulk fluid from the interface is called the double-layer thickness, 1/κ. This length is a function of the concentration and charges of the salts in solution, the temperature, and the permittivity of the solution:

κ² = (F²Σcz²) / (εε_oRT)

(2)

where ε = relative permittivity; ε_o = permittivity of free space = 8.854 × 10⁻¹⁴ C (V·cm); z = valency of each ion; c = concentration of each ion; R = 8.314 J (mol·K); T = temperature, K; and F = 9.65 × 10⁴ C/mol. The relative permittivity is a measure of the conductance of the pure bulk material relative to a vacuum. In the salt water in the beaker example, the pure bulk material would be water, which has a relative permittivity of about 80.

Not all of the ions in the diffuse layer are necessarily mobile. Sometimes the distinction is made between the location of the true interface, an intermediate interface called the Stern layer (5) where there are immobilized diffuse layer ions, and a surface of shear where the bulk fluid begins to move freely. The potential at the surface of shear is called the zeta potential. The only methods available to measure the zeta potential involve moving the surface relative to the bulk. Because the zeta potential is defined as the potential at the surface where the bulk fluid may move under shear, this is by definition the potential that is measured by these techniques (3).

The physical separation of charge represented allows externally applied electric field forces to act on the solution in the diffuse layer. There are two phenomena associated with the electric double layer that are relevant: electrophoresis when a particle is moved by an electric field relative to the bulk; and electroosmosis, sometimes called electroendosmosis, when bulk fluid migrates with respect to an immobilized charged surface.

Electrokinetics. The first mathematical description of electrophoresis balanced the electrical body force on the charge in the diffuse layer with the viscous forces in the diffuse layer that work against motion (6). Using this force balance, an equation for the velocity, V, of a particle in an electric field is

$$V = \mu E$$

(3)

where E = the electric field strength, V /cm; and $\mu = \frac{\epsilon\epsilon_o\zeta}{\eta}f(\kappa a)$ where ζ = zeta potential, mV; η = viscosity, g / (cm·s); and a = particle radius, cm. The function $f(\kappa a)$ is 1 when the particle has a much larger diameter than its double layer (high ionic strengths) and is 1.5 for very low ionic strength solutions where the double layer is much greater than the particle diameter (7). The behavior of a particle is complex when the double layer and particle diameter are close to the same size, which is often the case (8,9).

Electroosmotic flow is also dependent on the zeta potential at the immobilized surface and the strength of the electric field. For electroosmosis, the flow rate generated is

$$F = \frac{\epsilon\epsilon_o\zeta}{\eta}\pi r^2 E$$

(4)

where F = flow rate in mL /s and r = radius of flow channel in cm. Electroosmotic flow is generally minimized in polymer networks such as gels, but is important in open channels such as capillaries.

Generation of Heat in Electric Fields. One of the practical problems encountered in electrophoresis is the generation of heat from resistive dissipation of energy in the electrophoretic medium. The generation of heat (Joule heating) is given by

$$W = EI$$

(5)

where W = power in watts and I = current, $\mathcal{A}$. The current and the electric field strength are related by the conductivity of the electrophoretic medium by Ohm's law

$$E = \frac{I}{C}$$

(6)

where C = medium conductivity in $(\Omega\cdot\text{cm})^{-1}$. The higher the conductivity of the electrophoretic medium, the more difficult electrophoresis becomes because highly conductive solutions mean a lower field strength per current, and the heat load on the system increases as the current is squared. Electrophoresis is usually done in highly resistive media, using just enough salt to solubilize the compounds of interest. The addition of polymer matrices, such as gels, also serves to increase the resistivity of the media by increasing the viscosity.

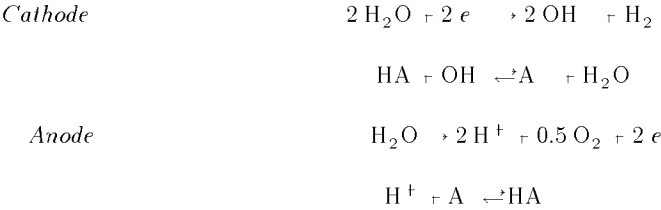
Heating in electrophoresis causes changes in the viscosity and density of the electrophoretic medium. High temperatures can also damage electrophoretic equipment by warping cooling blocks, melting plastic, or cracking glass plates. Poorly cooled electrophoretic systems have poor resolution. Usually bands are smeared or appear warped.

When fluids heat unevenly, the hot part of the fluid tends to rise with respect to the cooler part of the fluid because of differences in density. The flow is driven by gravity, and distorts resolution in electrophoretic separations.

The ability to remove heat from electrophoretic systems has severely limited the maximum capacity of these systems in terms of how large or thick the systems can be. Electrophoretic separations have been performed on space flights because the effect of gravity in outer space is small and mixing from heating is negligible. Whereas electrophoresis in outer space has been accomplished (10), the economics for a scaleable process have not (see SPACE PROCESSING).

The heating effect is the limiting factor for all electrophoretic separations. When heat is dissipated rapidly, as in capillary electrophoresis, rapid, high resolution separations are possible. For electrophoretic separations the higher the separating driving force, ie, the electric field strength, the better the resolution. This means that if a way to separate faster can be found, it should also be a more effective separation. This is the opposite of most other separation techniques.

Electrolysis Reactions. The electrodes in electrophoresis equipment are typically constructed from platinum wire, and sodium chloride generally carries the bulk of the current in any electrophoretic medium. This results in the reactions



That is, water is electrolyzed. The hydrogen gas produced at the cathode can be hazardous, especially because it is in the vicinity of an electrode that is also producing heat. For this reason, electrode chambers are usually open to the atmosphere so that gases can vent.

The other reactions at the electrodes produce acid (anode) and base (cathode) so that there is a possibility of a pH gradient throughout the electrophoresis medium unless the system is well buffered (see HYDROGEN ION ACTIVITY). Buffering must take the current load into account because the electrolysis reactions proceed at the rate of the current. Electrophoresis systems sometimes mix and recirculate the buffers from the individual electrode reservoirs to equalize the pH.

Modes of Electrophoretic Separations

Zone electrophoresis, isoelectric focusing, and isotachopheresis are all commonly practiced as analytical techniques. Zone electrophoresis is by far the most commonly practiced and there are several different zone electrophoresis methods and techniques available for different separation goals. Isoelectric focusing is also useful for separation, and possibly more useful for determining charge characteristics of sample proteins. Isotachopheresis takes advantage of the continuity of current across a medium to segregate a sample into contiguous zones of high purity. It is not useful as an analytical technique, but is applied as a potential preparative method.

Electrophoresis Equipment. Most electrophoresis equipment shares a basic design, a diagram of which is shown in Figure 1. Electrophoresis equipment usually consists of two buffer reservoirs, one for the anode and one for the cathode. Very often, electrophoretic buffers have a basic pH and sample compounds migrate to the anode. The equipment includes some sort of electrophoretic medium connecting the two reservoirs, such as a gel, paper, or capillary tubes, to which a sample is applied. A direct current power supply connects the two electrodes that are immersed in the buffer reservoirs. The power supply is usually interlocked to the electrophoresis equipment through a cover, isolating the medium and buffer reservoirs from human interaction because contact can be very dangerous. Small amounts of d-c current through the human heart can cause cardiac arrest or arrhythmia. An electric field is applied between the two electrode reservoirs, causing sample compounds to migrate through the medium and heat to be generated in the medium. The heat is usually removed by a chilled water bath or running tap water. At completion of the electrophoretic run the sample appears resolved into several different components along the length of the medium. In this respect, gel electrophoresis resembles thin-layer chromatography (see CHROMATOGRAPHY).

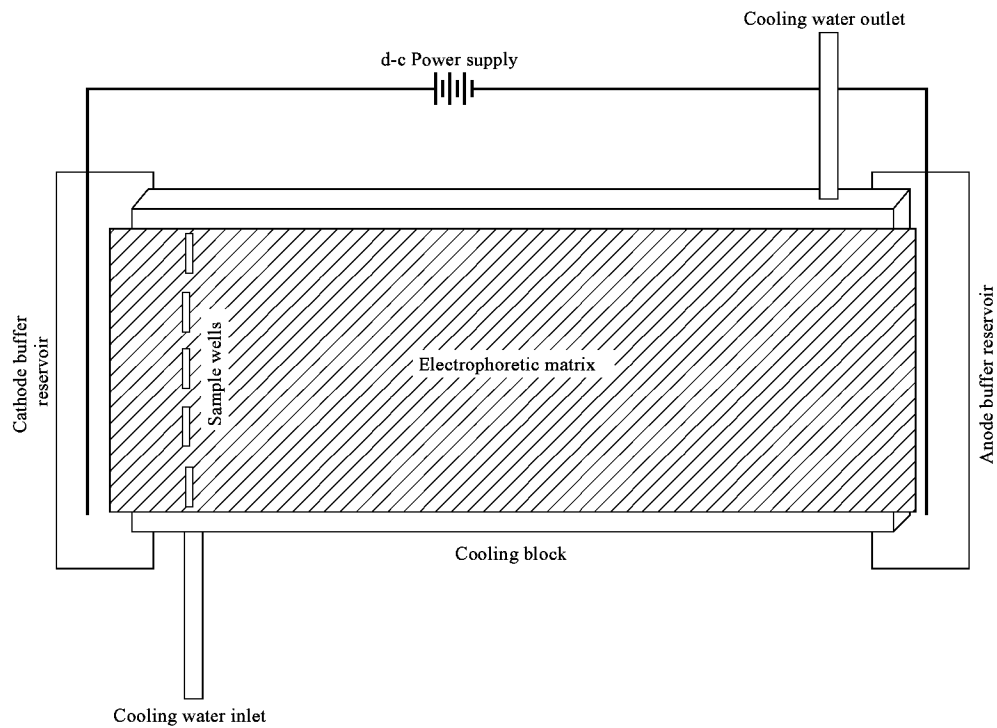


Fig. 1. Electrophoresis equipment.

Under certain conditions the sample is clearly visible throughout the process. Other times it is necessary to stain the matrix to visualize the components. In cases where a final staining procedure is required, a small amount of dye is often added to the sample before the analysis. The dye typically migrates faster than any sample component. The position of the dye in the matrix indicates the speed of the resolution of the components of the sample. Typically, the electrophoretic medium is discarded after use. Good resolution can be obtained from 1 to 20 hours, using applied voltages of 10 to 2000 V and currents of 5 to 100 mA.

Zone Electrophoresis. In zone electrophoresis multicomponent samples are applied to an electrophoretic medium, most commonly a gel, an electric field is applied, and after a predetermined length of time or after a certain level of power, current, or voltage has been applied, the electrophoretic medium is inspected for resolution of the sample components. A typical zone electrophoresis result is shown in Figure 2, which indicates how closely related molecules can be separated from each other and visualized on a gel. Each band represents a highly enriched substance, has essentially the same shape and width, and is separated from the other bands because each type of substance has a slightly different property. As shown, several samples are typically analyzed side by side on a gel. Bands that migrate the same distance in different sample mixtures represent the same substance and standards are usually run concurrently with samples that are to be compared. In zone electrophoresis, proteins or sample components are completely separated into discrete zones as they migrate through the media onto which they are applied.

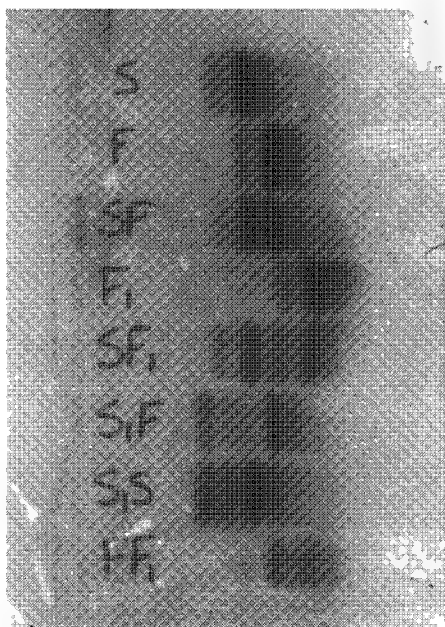


Fig. 2. Zone electrophoresis separation where S, F, S₁, and F₁ are different materials.

The use of standards with samples makes zone electrophoresis particularly useful as an analytical tool. However, when samples cannot be analyzed on the same gel, differences in the experimental conditions from experiment to experiment make direct comparison more difficult. To make comparisons from experiment to experiment, a relative mobility, R_f , is often measured by measuring the distance a component travels down the gel compared to some reference or standard component.

$$R_f = \frac{\text{distance migrated by component}}{\text{distance migrated by reference}} \tag{7}$$

Disc Electrophoresis. Resolution in zone electrophoresis depends critically on getting sample components to migrate in a focused band, thus some techniques are employed to concentrate the sample as it migrates through the gel. The most common technique is referred to as discontinuous pH or disc electrophoresis. Disc electrophoresis employs a two-gel system, where the properties of the two gels are different.

Disc electrophoresis was first introduced in the early 1960s (11–13) as various techniques using polyacrylamide gels were being explored and designed. Original work employed several buffer systems and different polyacrylamide gels in order to first concentrate and then separate compounds (14).

The way proteins behave in disc electrophoresis systems depends primarily on differences in the pH in the two gels. The pH of the buffers in both the second gel (the separating gel) and the electrode reservoir are similar, whereas the buffers of the first gel (the stacking gel) and the sample itself are of a lower pH. This difference in pH allows for different sample–gel interactions. The stacking gel is only about 5 cm, including distance for the sample wells, and is stacked on top of the separating gel which is about 20 cm in length. As with most electrophoretic methods, a current is applied and the molecules in the sample wells begin to migrate anodally. Here, the migration is electrochemically different because the buffer in the upper reservoir chamber has a higher pH than that of the samples and stacking gel. This difference in pH allows the molecules in the samples to migrate rapidly through the stacking gel. When the sample compounds enter the separating gel, movement is slowed because of the pH change. This focuses the molecules into narrow bands and allows more bands to be resolved from one another.

Another difference between other types of electrophoresis and disc electrophoresis is that the molecules in a sample do not start to significantly separate until entering the separating gel. A discontinuous gel system may be used with almost any type of zone electrophoresis application.

Native Zone Electrophoresis. In some cases, good resolution between sample species can be obtained with little or no sample pretreatment. In these cases, the gels are said to be native gels. In this method, the charge on the individual sample component is primarily responsible for its differential migration. The relative mobility, R_f , measured is proportional to the charge when the pore size in the electrophoretic medium is large compared to the component. When the pore size and molecular size are about the same, the size of the molecule becomes important in the electrophoretic mobility. Because of this ambiguity, the absolute meaning of R_f from this technique is useful primarily for component identification and comparison, and not for estimating the properties of a molecule.

When separating proteins, native zone electrophoresis leaves the component proteins in folded, globular states, with all subunits intact. This compact size makes for a faster running molecule. Native electrophoresis is useful for resolving components which may differ by a small size and or a small charge. For example, this method was used to isolate genetic variants of some proteins in cow's milk (15) which differ by one charge unit and less than 0.1% of their total molecular weight.

Reduced SDS Electrophoresis. The combination of sodium dodecyl sulfate (SDS) [151-21-3], $\text{CH}_3(\text{CH}_2)_{10}\text{CH}_2\text{OSO}_3\text{Na}$, also known as lauryl sulfate, treatment of samples and polyacrylamide gel electrophoresis was first described in the late 1960s (14,16). SDS is an ionic surfactant which solubilizes and denatures proteins (see SURFACTANTS). The surfactant coats a protein through hydrophobic interactions with the polypeptide backbone, effectively separating most proteins into their polypeptide subunits. The majority of proteins to which SDS binds then unfold into linear molecules having a similar surface potential. Nonreduced proteins bind approximately 0.9–1.0 grams of SDS per gram of protein (17).

SDS–polyacrylamide gel electrophoresis (SDS–page) allows separation of molecules strictly on the basis of size, ie, molecular weight. When SDS-treated samples migrate into a gel and are electrophoresed, the principal difference is size or length. Smaller molecules travel through the matrix more quickly than those that are larger. The rate at which molecules migrate through a polyacrylamide gel is inversely linear with the logarithm of their molecular weight. Thus denatured samples can be analyzed alongside standards of known molecular weight to aid in the interpretation of a substance's physical size.

Other dissociating agents may be used to further break down a protein. Urea [57-13-6] may be used to disrupt hydrogen bonds. When urea is the only dissociating agent added (no SDS), a protein's intrinsic charge is not affected and separation on the basis of size and charge may be achieved. If a protein contains internal disulfide bonds, a thiol reagent such as β -mercaptoethanol [60-24-2] must be used in order to reduce the sample and break the disulfide bonds. Proteins having reduced disulfide bonds bind approximately 1.4 grams of SDS per gram of protein, compared to about 1 g/g for nonreduced. Typically, both a reduced and nonreduced sample are run in order to evaluate band differences.

Pulsed Field Gel Electrophoresis. Pulsed field gel electrophoresis is a technique that was developed to separate large pieces of DNA in agarose gels. DNA had previously been separated by conventional size sieving electrophoretic techniques. The resolving power of this technique is inversely proportional to the log of the DNA molecular weight. Thus as the molecular weight of the DNA increases, the resolution decreases. In pulsed field electrophoresis, the direction of the field is intermittently changed, either forward and backward or from side to side. A small molecule notices no real difference in its electrophoresis because it can completely reorient in each field direction. However, the redirectioning of the electric field causes larger molecules to travel in a zigzag pattern, putting kinks into the length of the molecule. The longer the molecule, the more kinks in its length, and the slower it travels down the length of the gel. The larger molecule finds itself traveling backward along some sections of its length with respect to the direction of the electric field. This has allowed resolution of megabase size strands of DNA, and has made the Human Genome Project feasible (18–23).

Isoelectric Focusing. Isoelectric focusing is a technique used for protein separation, by driving proteins to a pH where they have no mobility. Resolution depends on the slope of a pH gradient that can be achieved in a gel.

Ampholytes or Zwitterions. An ampholyte is a molecule that can be either positively or negatively charged, depending on the pH. These molecules are also called zwitterions. All amino acids (qv) and proteins are ampholytes, or amphoteric. Not only does the sign of the charge of an ampholyte change with pH, but the magnitude of the charge can also vary. The charge on a protein, for example, may vary from +10 or more at low pH, to 10 or more at high pH. For a protein, the pH at which its mobility is zero is called the isoelectric point.

A special class of ampholytes has been synthesized for the purpose of isoelectric focusing (24). These ampholytes have an amino end and a carboxyl end that are separated by varying numbers of methylene groups. The further apart the amino and carboxyl groups, the less one affects the ionization of the other; thus a different isoelectric point (pI) is established for each molecule. These ampholytes, which may be added to an electrophoretic medium, migrate in one direction or another, under the influence of an applied electric field, until they reach a zone in which the pH is the same as that ampholyte's isoelectric point. The ampholyte molecules buffer themselves and establish the local pH as they migrate through the gel. As the ampholytes reach an isoelectric pH, they establish a stationary spatial pH gradient in the electrophoretic medium.

Isoelectric Focusing. Isoelectric focusing (ief) is an electrophoretic technique in which amphoteric samples are separated according to their isoelectric points along a continuous pH gradient. Ief analyses are carried out in various matrices: in acrylamide, agarose, and capillaries. The agarose or acrylamide gels that are used must be prepared with carrier ampholytes bracketing a specific pH range. After some time, the ampholytes separate and there is a pH gradient which covers the range of all the ampholytes' isoelectric points. Initial research on ief (25) was primarily directed toward evaluating the properties of synthetic ampholytes in solution. Later work refined the technique to apply gel matrices (12) and provided the basis for ief methodologies.

Problems associated with gel-to-gel variability have been rectified with the advancement of ampholyte mixtures. One commonly used mixture of ampholytes, called Immoblines, is an improved ampholyte mixture that produces no gradient drift or unequal pH gradient (26) and can be used in a gel matrix reproducibly from one day to the next.

Because protein samples are actually ampholytes, when samples are loaded onto the gel and a current is applied, the compounds migrate through the gel until they come to their isoelectric point where they reach a steady state. This technique measures an intrinsic physicochemical parameter of the protein, the pI, and therefore does not depend on the mode of sample application. The highest sample load of any electrophoretic technique may be used, however, sample load affects the final position of a component band if the load is extremely high, ie, high enough to titrate the gradient ampholytes or distort the local electric field.

Isoelectric focusing takes a long (from ca 3 to 30 h) time to complete because sample compounds move more and more slowly as they approach the pH in the gel that corresponds to their isoelectric points. Because the gradient ampholytes and the samples stop where they have no mobility, the resistivity of the system increases dramatically toward the end of the experiment, and the current decreases dramatically. For this reason, isoelectric focusing is usually run with constant voltage. Constant current application can lead to overheating of the system.

Some forms of agarose are specifically designed to work with large (mol wt >500,000) molecules (27,28). The types of samples for which the agarose ief system are utilized are larger plasma proteins such as immunoglobulins, tissues, and tumors.

Another form of ief is a method called direct tissue isoelectric focusing (dtif) (29) where isoelectric focusing in agarose is used to evaluate tissues. The tissue to be analyzed is placed directly onto the gel. Using the tissue itself and not tissue extracts has advanced the study of proteins that are difficult to extract from tissue, or are damaged by the extraction procedure. Dtif is an important advancement in the area of sample handling and application where direct application of a solid to a gel matrix may actually enhance resolution.

Isotachophoresis. Isotachophoresis takes advantage of the fact that electroneutrality must be maintained in an electrophoretic system in order to support an electric field. If a current passes through a medium, that current must be constant from one electrode to the other, regardless of the local ion concentration or mobility; ie, dilute ions must move faster to keep up with a zone of more concentrated ions. Electric fields compensate for this because the electric field strength does not have to be constant along the length of the medium. The electric field strength is lowest where the ions are most concentrated and most mobile. Isotachophoresis takes advantage of this phenomenon by lining up the ions of interest, fastest (most mobile) to slowest. This is a highly specialized technique that requires detailed knowledge of the properties of the sample to be separated, and is generally not applicable to analytical separations.

An electrophoretic medium, such as a gel or a capillary tube, is cast or filled with an electrolyte that has a higher mobility than any components in the mixture of interest. This electrolyte, called the leading electrolyte, also fills the anode buffer reservoir. A sample is applied on top of the leading electrolyte, and another electrolyte with lower mobility than the sample fills the cathode reservoir. This essentially envelopes the sample. As the electric field is applied, the leading electrolyte moves rapidly toward the anode. The highest mobility component in the sample is drawn toward the lead electrolyte to fill in the conductivity gap left by the quickly migrating lead electrolyte. The electric field driving the first ion in the sample must increase to allow the sample ion to follow at the same speed as the leading electrolyte. The current through the whole media is constant, so the electric field strength increases because the resistivity increases. The resistivity increases because the first ion in the sample to follow the lead electrolyte dilutes until the electric field is balanced to give the two zones the same speed. Progressively, each ion in the sample follows at lower conductivity, just behind an ion of higher mobility and just ahead of an ion of lower mobility. Finally, the "trailing" electrolyte, of lowest mobility, terminates the separation.

This separation technique has been employed primarily for preparative types of separations because detailed knowledge of the properties of the sample is required. Also, because this separation results in discrete zones of sample ions which are virtually pure, it makes sense to use this technique when the sample size is large. This technique is ineffective when the levels of impurities are small with respect to the target compound; small amounts of sample ions do not form zones well and tend to mix with the target compound. Information on this technique is available (30).

Electrophoretic Materials and Matrices

Various support media may be employed in electrophoretic techniques. Separation on agarose, acrylamide, and paper is influenced not only by electrophoretic mobility, but also by sieving of the samples through the polymer mesh. The finer the weave of selected matrix, the slower a molecule travels. Therefore, molecular weight or molecular length, as well as charge, can influence the rate of migration.

In addition to polymeric support media, capillaries and flowing buffers have been used as support media for electrophoresis. Although these are not used as frequently, there are definite advantages for certain types of samples and applications.

Agarose Electrophoresis. Agarose is produced from the processing of red seaweed. When agar is extracted from the seaweed it is in two components, agarpectin and agarose. The agarose portion is nearly uncharged and is therefore the portion desirable for use as an electrophoretic matrix. Charge on the agarose leads to electroosmotic flow through the gel. The chemical composition of agarose is alternatively repeated residues of 1,3-linked-β-D-galactopyranose and 1,4-linked-3,6-anhydro-α-L-galactopyranose (31).

To prepare a gel for electrophoresis a combination of agarose and buffer is heated until the agarose solid is dissolved and boiling. The solution is cooled sufficiently and then poured into a warmed gel casting apparatus which forms the shape of the gel as it cools. After the cooled solution is poured into the casting apparatus, it is allowed to gel and can then be used in an agarose electrophoresis method. Because the composition of agarose is a network of residues that hold water molecules, the extra agarose solution may be stored and used at a later time. The concentration of agarose mixtures typically varies between 0.5 and 2% in a weight-to-volume ratio, although more extreme concentrations have been evaluated. The varying concentrations depend on the desired application. The higher the concentration of agarose, the smaller the pore size.

The use of agarose as an electrophoretic method is widespread (32–35). An example of its use is in the evaluation and typing of DNA both in forensics (see FORENSIC CHEMISTRY) and to study heritable diseases (36). Agarose electrophoresis is combined with other analytical tools such as Southern blotting, polymerase chain reaction, and fluorescence. The advantages of agarose electrophoresis are that it requires no additives or cross-linkers for polymerization, it is not hazardous, low concentration gels are relatively sturdy, it is inexpensive, and it can be combined with many other analytical methods.

Polyacrylamide Electrophoresis. Polyacrylamide gels are synthesized through the combination of acrylamide [79-60-1] (qv), CH₂=CHCONH₂, monomer and a cross-linking comonomer (see ACRYLAMIDE POLYMERS). Typically, the cross-linking comonomer of choice is N,N'-methylenebisacrylamide [110-26-9] (bisacrylamide), (CH₂CHCONH)₂, although others are available, such as ethylenediacrylate (EDA) and N,N'-diallyltartardiamide (DATD) [58477-85-3] (37). The cross-linking of polymerized monomer with the comonomer is what controls the pore size of the gel polymer mesh. This level of pore size control makes polyacrylamide gel electrophoresis an effective analytical tool.

The most commonly used combination of chemicals to produce a polyacrylamide gel is acrylamide, bisacrylamide, buffer, ammonium persulfate, and tetramethylenediamine (TEMED). TEMED and ammonium persulfate are catalysts to the polymerization reaction. The TEMED causes the persulfate to produce free radicals, causing polymerization. Because this is a free-radical driven reaction, the mixture of reagents must be degassed before it is used. The mixture polymerizes quickly after TEMED addition, so it should be poured into the gel-casting apparatus as quickly as possible. Once the gel is poured into a prepared form, a comb can be applied to the top portion of the gel before polymerization occurs. This comb sets small indentations permanently into the top portion of the gel which can be used to load samples. If the comb is used, samples are then typically mixed with a heavier solution, such as glycerol, before the sample is applied to the gel, to prevent the sample from dispersing into the reservoir buffer.

Maximum resolution of proteins is the main consideration when determining the acrylamide concentration of a gel. Two parameters define the composition of gel: the wt % of total monomer, % T (acrylamide plus bisacrylamide in grams per 100 mL), and the proportion by weight of monomer that is the cross-linking agent, % C (bisacrylamide) (38). Gels having concentrations between 5 and 15% are typically used to achieve the most desirable separation of the components of interest. Unlike the agarose matrix, once the mixture polymerizes, it cannot be reused.

Polyacrylamide gel electrophoresis is one of the most commonly used electrophoretic methods. Analytical uses of this technique center around protein characterization, for example, purity, size, or molecular weight, and composition of a protein. Polyacrylamide gels can be used in both reduced and nonreduced systems as well as in combination with discontinuous and ief systems (39).

An example of the use of polyacrylamide gels in an ief system in combination with immunoblotting is given in Reference 40 where this method is used to detect low quantities of group specific component (GC) subtypes. Another example of polyacrylamide in combination with ief is given in Reference 41 where the technique is used as a screening tool for inheritance of a certain polymorphic protein.

Both the ease of use of this method for characterization of proteins and nucleic acids, and the ability to analyze many samples simultaneously for comparative purposes, have led to the prevalence of this technique. The drawbacks of a polyacrylamide matrix is that acrylamide is a neurotoxin, the reagents must be combined extremely carefully, and the gels are not as pliable as most agarose gels.

Paper Electrophoresis. Paper (qv) as an electrophoretic matrix was employed in some of the first electrophoretic techniques developed to separate compounds. Paper is easier than a gel matrix because the paper matrix requires no preparation. Besides being easy to obtain, paper is a good medium because it does not contain many of the charges that interfere with the separation of different compounds. Two types of paper employed in this type of electrophoresis are Whatman 3 MM (0.3 mm) and Whatman No. 1 (0.17 mm).

In paper electrophoresis, the sample is placed directly onto chromatographic or filter paper and then exposed to a buffer solution at each end and an electric field is applied. As in most electrophoretic techniques, charged dyes are combined with samples and standards to see the progress of the electrophoresis. The movement of samples on paper is best when the current flow is parallel to the fiber axis in the paper. The paper has high resistance so voltages are typically much higher than in agarose and polyacrylamide matrices. Like agarose and polyacrylamide matrices, paper is combined with other analytical tools to enhance separation and identification of sample components. For example, paper electrophoresis has been combined with chromatography (42).

The difference between paper electrophoresis and paper chromatography is that electrophoresis separates by charge whereas chromatography

separates by polarity. This combined technique was used to evaluate polymorphisms of the hemoglobin molecule, ie, normal A-type versus the sickle cell S-type. It has been called peptide fingerprinting. This method was later modified (43) to further evaluate peptides and is one technique known as peptide mapping. The peptide mapping technique also uses high voltages to obtain the desired resolution.

Some advantages of paper in electrophoresis are that paper is readily available, easy to handle, and new methodologies can be developed rapidly. The disadvantages of paper electrophoresis are that the porosity of paper cannot be controlled, the technique is not very sensitive, and it is not easily reproducible.

Capillary Electrophoresis. Capillaries were first applied as a support medium for electrophoresis in the early 1980s (44,45). The glass capillaries used are typically 20 to 200 μm in diameter (46), may be filled with buffer or gel, and are frequently coated on the inside. Capillaries are used because of the high surface-to-volume ratio which allows high voltages without heating effects. The only limitations associated with capillaries are limits of detection and clearance of sample components.

Limits of detection become a problem in capillary electrophoresis because the amounts of analyte that can be loaded into a capillary are extremely small. In a 20 μm capillary, for example, there is 0.03 μL cm capillary length. This is 1/100 to 1/1000 of the volume typically loaded onto polyacrylamide or agarose gels. For trace analysis, a very small number of molecules may actually exist in the capillary after loading. To detect these small amounts of components, some on-line detectors have been developed which use conductivity, laser Doppler effects, or narrowly focused lasers (qv) to detect either absorbance or fluorescence (47,48). The conductivity detector claims detection limits down to 10⁻¹² moles. The laser absorbance detector has been used to measure some of the components in a single human cell (see TRACE AND RESIDUE ANALYSIS).

Clearance of sample components from a capillary is a problem that has not been as well resolved as detection. Usually, capillary electrophoresis is conducted in a glass capillary that is coated on its inner surface. Some applications call for gel-filled capillaries. Capillaries are difficult to coat or fill with gel, and are too expensive to discard after each use, as is done when using a standard gel. Therefore, all the sample has to be cleared from the capillary before the capillary is used for another sample, much like analytical chromatography. This can be a problem when capillaries are sometimes more than 100 cm long. Another problem arises when analytes absorb to the glass (or coated) walls in the capillary, a frequent occurrence. This latter problem has prevented capillary electrophoresis from becoming a practical method for protein analysis. In coated capillaries, electroosmosis along the capillary walls sometimes adds velocity to the analytes, and clearance is less of an issue. A great deal of research has focused on choosing and optimizing capillary coatings that resist protein fouling and absorb sufficient charge from buffer ions to give electroosmotic flow.

Electroosmotic flow in a capillary also makes it possible to analyze both cations and anions in the same sample. The only requirement is that the electroosmotic flow downstream is of a greater magnitude than electrophoresis of the oppositely charged ions upstream. Electroosmosis is the preferred method of generating flow in the capillary, because the variation in the flow profile occurs within a fraction of κr from the wall (49). When electroosmosis is used for sample injection, differing amounts of analyte can be found between the sample in the capillary and the uninjected sample, because of different electrophoretic mobilities of analytes (50). Two other methods of generating flow are with gravity or with a pump.

Capillary electrophoresis is a commercially available technique, and has been integrated with most automated lab equipment such as autosamplers, computer peak analysis (the charts generated are called electropherograms), temperature control, and recirculating buffers. Capillary electrophoresis separations are rapid (minutes), require high voltage sources (10,000 V), and a small amount of automation, particularly in sample application, to be feasible. The use of capillaries as a support medium for electrophoresis is advantageous because it avoids the effects of heating that occur in a gel, can be very rapid, and produces a chart recording rather than a stained gel for archiving.

Free-Flow Electrophoresis. Free-flow electrophoresis is the most common technique for scaling up electrophoresis for commercial application. In this technique, sample compounds are injected into a curtain of buffer which flows between two flat plates, with electrodes parallel to the flow at each end. The electric field is then applied perpendicularly to the flow direction, so that as compounds flow down between the electrodes they separate horizontally and exit the flow field at different locations. The main challenge for this technique is stabilizing the flow to both heating and electroosmotic forces. Sometimes this is done by dividing the flow curtain into cells using semipermeable membranes, which allow proteins and other sample compounds to migrate from chamber to chamber, but restrict flow (see MEMBRANE TECHNOLOGY). Another method is to apply a very low electric field so that little heating and electroosmosis occur, but then to recycle the material through coolers. The material is then sent back through the separating cell (51–53).

Most electrophoretic methods have been tried in a free-flow format, including isoelectric focusing, native zone electrophoresis, and isotachopheresis. Most free-flow electrophoresis equipment has very low (ca 1 g (L·h)) capacity, and resolution is reduced by heating and electroosmotic considerations.

Detection Techniques

Most sample components analyzed with electrophoretic techniques are invisible to the naked eye. Thus methods have been developed to visualize and quantify separated compounds. These techniques most commonly involve chemically fixing and then staining the compounds in the gel. Other detection techniques can sometimes yield more information, such as detection using antibodies to specific compounds, which gives positive identification of a sample component either by immunoelectrophoretic or blotting techniques, or enhanced detection by combining two different electrophoresis methods in two-dimensional electrophoretic techniques.

Chemical Staining. Staining techniques that help to visualize banding patterns resulting from electrophoresis vary (54–56). The size and type of the compound as well as the electrophoretic matrix dictate and often limit the variety of stains that can be used. Molecules can be lost during the staining process, so most staining procedures incorporate a "fixing" step either before or in conjunction with staining. With paper electrophoresis, proteins are fixed to the medium by drying the paper. With agarose and acrylamide gels, the fixation of proteins requires additives. Acidic solutions, such as trichloroacetic acid [76-03-9], are commonly used to fix proteins of all sizes. Once proteins are fixed, they can be stained without loss of the separated components.

Amido black is a commonly used stain, but it is not very sensitive. It is often used to visualize concentrated proteins or components that are readily accessible to dyes such as proteins that have been transferred from a gel to nitrocellulose paper. Two of the more sensitive and more frequently used stains are Coomassie Brilliant Blue (R250 and G250) and silver stains. Because these stains interact differently with a variety of protein molecules, optimization of the fixative and staining solutions is necessary. The Coomassie stains are approximately five times more sensitive than amido black and are appropriate for both agarose and polyacrylamide gels. The silver stain is approximately 100 times more sensitive than Coomassie and is typically used for polyacrylamide gels.

A silver stain is used when proteins exist in a very small quantity or when analysis of as many bands as possible created by separation techniques is desired. One positive application of silver stain is its sensitivity. A drawback of the silver stain, however, is that it is more complex and often requires more troubleshooting to obtain the desired results.

To quantitate proteins from staining, a densitometer aided by computer software is used to evaluate band areas of samples compared to band areas of a standard curve. Amido black, Coomassie Brilliant Blue, and silver stains are all applicable for use in quantification of proteins.

Another method to visualize and identify separation products on a gel is through radioactivity. If a sample is radioactive, the bands that form through migration of a sample are subsequently radioactive. This type of gel may be placed against an x-ray until the radiation makes a mirror image of the banding pattern on the x-ray. The x-ray is then developed and the resulting autoradiograph displays the bands.

As an alternative to radiation, a stain such as ethidium bromide is used to visualize DNA. The ethidium may be incorporated into the structure of DNA either before or after electrophoresis. The gel is then visualized under a fluorescent lamp.

Immunoelectrophoretic Techniques. The technique of gel electrophoresis has been successfully combined with immunological techniques in order to further evaluate molecules. Specifically, the concept of double immunodiffusion as described in 1948 (57) and that of single-radial immunodiffusion described in 1963 (58) have been further developed for use with electrophoresis in both the clinical and research setting.

The double-immunodiffusion technique, often referred to as the Ouchterlony technique, uses an agarose gel as the matrix. Holes are made in the agarose where either sample or antisera is placed. The two solutions are allowed to diffuse into the matrix for a predetermined time. If there is a reaction

between the antigen in the sample and the antibody, a precipitate is formed. The Ag–Ab reaction line can be visualized after a stain is used (Fig. 3a). Similarly, the single-radial immunodiffusion (Fig. 3b) technique is a reaction of equilibrium between antigen and antibody. The difference is that the antigen is added to a warmed agarose solution before it gels. Holes are cut in the gel as for the Ouchterlony technique where samples are placed. The sample diffuses into the gel and, if reactive with the antibody, forms an area of precipitation around the hole that is visualized after staining.

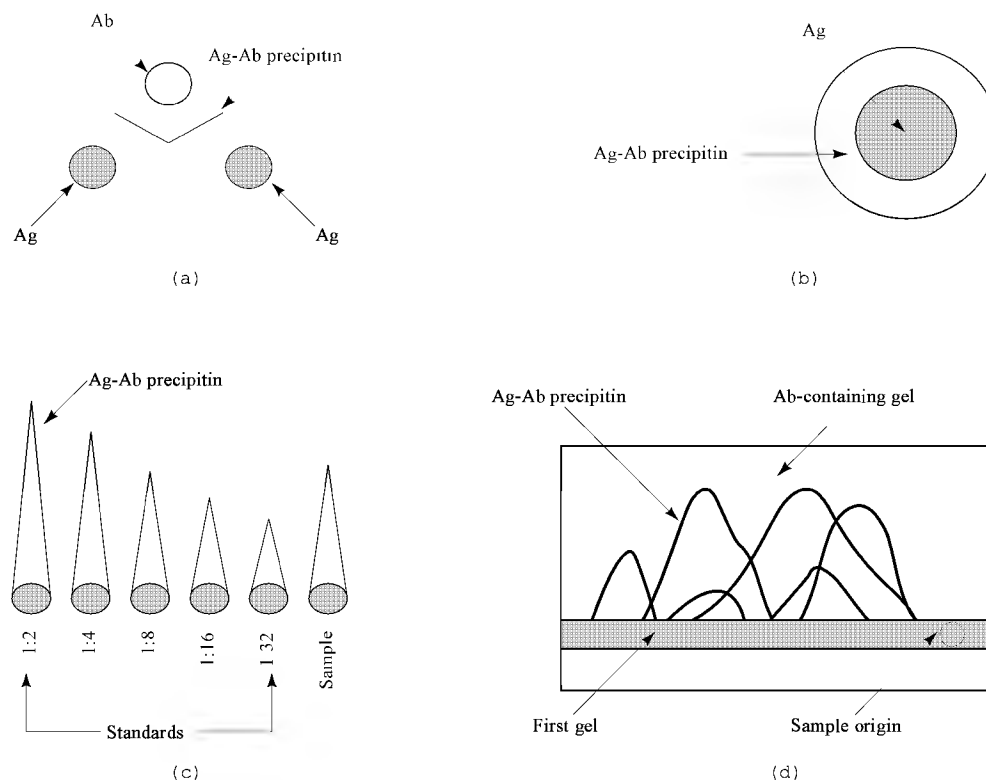


Fig. 3. Immunological reactions, where Ag is antigen and Ab is antibody, for detection in electrophoresis: (a) Ouchterlony technique; (b) single-radial diffusion; (c) rocket immunoelectrophoresis; and (d) crossed immunoelectrophoresis.

The Mancini and Ouchterlony techniques are the basis of the techniques employed in immunoelectrophoresis. A technique referred to as rockets (59) is named as such because of the appearance of a rocket-shaped antigen–antibody precipitin formed after an antigenic sample is electrophoresed through a gel-containing antibody (Fig. 3c).

Another frequently used method of immunoelectrophoresis is a technique known as crossed immunoelectrophoresis (Fig. 3d) (60). A sample is first run vertically through an agarose gel for a predetermined time. Secondly, the gel area where the sample was electrophoresed is typically cut out and placed horizontally into a similarly sized area of an antibody-containing gel. As an electrical current is applied to the second gel system, the sample in question electrophoreses through the gel and forms an antigen–antibody precipitin over an area that varies from small to large, depending on the banding pattern of electrophoresis from the first gel system.

At a somewhat more basic level, both agarose and acrylamide gel systems have been used for direct immunofixation. In these gels, samples are electrophoresed and then immunofixed by either using strips of cellulose acetate soaked in an antibody or the antibody is placed directly over the sample area of the gel.

All of these techniques are most often, but not exclusively, used in the clinical setting in order to diagnose abnormalities or to evaluate inheritance patterns of polymorphic proteins. Many applications of these techniques exist (61–65).

Two-Dimensional Electrophoresis. Two-dimensional (2D) electrophoresis is unique, offering an analytical method that is both reproducible and sensitive. It is referred to as 2D because it employs two different methods of electrophoresis, in two different dimensions, to produce one result. Each method separates the sample compounds based on different properties of each compound. The combination of the two methods gives better resolution of the compounds in the sample than could be achieved with either method alone. For example, each method alone may separate up to 100 components of a sample, whereas together they may separate up to 10,000 components.

A pair of electrophoretic techniques commonly employed in 2D analyses are isoelectric focusing (ief) and SDS–polyacrylamide gel electrophoresis (SDS–page). Ief separates sample compounds according to isoelectric point, whereas SDS–page separates the compounds by molecular weight. A 2D analytical technique using ief and SDS–page to separate total protein results in a gel having bands or spots in a random pattern (66). Each spot represents a unique component of a sample. A single charge difference in a component can be identified on the gel by a unique spot. This property of 2D electrophoresis, which allows identification of identical proteins that differ by one charge difference, has made it an invaluable technique for the molecular genetic community. Software is available to identify each spot on a gel and compare abnormalities in samples such as human blood (67), *Escherichia coli*, and yeast proteins.

Blotting Techniques. Problems encountered when trying to analyze resolved components of a sample mixture on a gel, with techniques such as direct immunofixation or application of a ligand, can be circumvented with the use of blotting techniques followed by staining or autoradiography. It is the inability of some compounds, such as antibodies or ligands, to enter the gel matrix of a specific gel system which leads to the development of various blotting techniques which have become widespread since the late 1970s. The nucleic acid and protein blotting techniques have become useful because these combine electrophoretic analyses and sensitive immunological tools. A blotting technique involves the transfer of nucleic acids or proteins, immediately after being separated by electrophoresis, from the gel matrix to another matrix. Typically, the other matrix is nitrocellulose paper, nylon, or other high affinity membrane and the mode of transfer is electrotransfer for proteins and capillary transfer for nucleic acids. Once nucleic acids or proteins are transferred, further analyses can be performed. On nitrocellulose paper or nylon, nucleic acids and proteins are more accessible than in the original gel matrix. This second matrix is then treated with a ligand to identify a specific component of a sample.

For example, the technique of Southern blotting was developed (68) for use with agarose gel electrophoresis of DNA fragments. Southern blots are designed to detect specific sequences of DNA. After electrophoresis is complete, the DNA is denatured and the single stranded DNA transferred to the specially prepared nitrocellulose paper. The nitrocellulose is then incubated with radioactive RNA or DNA complementary to those DNA sequences of interest. After the nitrocellulose has been sufficiently incubated with the radioactive complementary DNA, autoradiography is used to identify the fragments of interest.

The northern blotting technique is similar to the Southern blotting technique with the exception that northern blots detect specific sequences of RNA, not DNA.

The technique of immunoblotting, often referred to as western blotting, is frequently used for a variety of applications where protein concentrations

are low and staining of the electrophoretic matrix does not produce adequate resolution. In these instances, as with Southern blotting, proteins are transferred to a second matrix like nitrocellulose or nylon and are then treated with antisera specific to a desired protein (69,70).

Blotting techniques may be used in a variety and combination of electrophoretic systems which makes their use widespread and convenient when protein concentrations are minimal and agarose or polyacrylamide is the matrix choice.

BIBLIOGRAPHY

1. J. S. Newman, *Electrochemical Systems*, Prentice Hall, Inc., Englewood Cliffs, N.J. 1973.
2. R. A. Robinson and R. H. Stokes, *Electrolyte Solutions*, Butterworths Scientific Publications, London, 1959.
3. R. J. Hunter, in R. H. Ottewill and R. L. Rowell, eds., *Zeta Potential in Colloid Science, Principles and Applications*, Academic Press, Inc., New York, 1981.
4. S. R. Rudge and P. Todd, in M. R. Ladisch, R. C. Willson, C. C. Painton, and S. E. Builder, eds., *Protein Purification, From Molecular Mechanisms to Large-Scale Processes*, ACS Symposium Series 427, Washington, D.C. 1990, p. 244.
5. O. Stern, *Zeitschr. Elektrochem.* **30**, 508 (1924).
6. M. von Smoluchowski, *Bull. Akad. Sci. Cracovie, Classe Sci. Math. Natur.* **1**, 182 (1903).
7. E. Hückel, *Physik. Zeitschr.* **24**, 204 (1924).
8. D. C. Henry, *Proc. Roy. Soc. (London) Ser. A* **133**, 106 (1931).
9. R. W. O'Brien and L. R. White, *J. Chem. Soc. Faraday II* **74**, 1606 (1978).
10. K. Hannig, H. Wirth, and E. Schoen, *Apollo-Soyuz Test Project Summary Science Report*, Vol. 1, NASA SP-412, Washington, D.C., 1977, p. 335.
11. B. J. Davis, *Ann. N.Y. Acad. Sci.* **121**, 404 (1964).
12. L. Ornstein, *Ann. N.Y. Acad. Sci.* **121**, 321 (1964).
13. C. Schafer-Nielsen, in M. J. Dunn, ed., *Gel Electrophoresis of Proteins*, Wright, Bristol, UK, 1986, p. 3.
14. U. K. Laemmli, *Nature* **227**, 680 (1970).
15. R. Aschaffenburg and J. Drewry, *Nature* **176**, 218 (1955).
16. J. V. Maizel, in K. Habel and N. P. Salzman, eds., *Fundamental Techniques in Virology*, Academic Press, Inc., New York, 1969, p. 334.
17. R. Pitt-Rivers and F. S. A. Impiombato, *Biochem. J.* **109**, 825 (1968).
18. D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984).
19. G. F. Carle, M. Frank, and M. V. Olson, *Science* **232**, 65 (1986).
20. G. F. Carle and M. V. Olson, *Nucleic Acids Research* **12**(14), 5647 (1984).
21. G. Chu, D. Vollrath, and R. W. Davis, *Science* **234**, 1582 (1986).
22. F. S. Collins and co-workers, *Science* **235**, 1046 (1986).
23. S. K. Lawrence, C. L. Smith, R. Srinastava, C. R. Cantor, and S. M. Weissman, *Science* **235**, 1387 (1986).
24. O. Vesterberg, *Acta Chem. Scand.* **23**, 2653 (1969).
25. H. Svensson, *Acta Chem. Scand.* **15**, 325 (1961).
26. A. Chrambach, *The Practice of Quantitative Gel Electrophoresis*, VCH Publishers, Deerfield Beach, Fla., 1985, p. 141.
27. A. Rosen, K. Ek, and P. Aman, *J. Immunol. Methods* **28**, 1 (1979).
28. C. A. Saravis, M. O'Brien, and N. Zamcheck, *J. Immunol. Methods* **29**, 97 (1979).
29. C. A. Saravis and N. Zamcheck, *J. Immunol. Methods* **29**, 91 (1979).
30. F. M. Everaerts, F. E. P. Mikkers, and Th. P. E. M. Verheggen, *Sep. Purif. Meth.* **6**, 287 (1977).
31. C. Araki, *Bull. Chem Soc. Jpn.* **29**, 543–544 (1956).
32. C. A. Alper, in W. J. Williams, E. Beutler, A. J. Erslev, and M. A. Lichtman, eds., *Hematology*, 3rd ed., McGraw-Hill Books, Inc., New York, 1983, p. 1516.
33. E. S. Lander, *Nature* **339**, 501 (1989).
34. T. Maniatis, E. F. Fritsch, and J. Sambrook, *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1983, pp. 150–171.
35. R. F. Boyer, *Modern Experimental Biochemistry*, The Benjamin Cummings Publishing Co., Inc., 1986, p. 219.
36. S. Wood and S. Langlois, *J. Chromatog.* **569**, 421 (1991).
37. Ref. 26, p. 64.
38. D. E. Garfin, *Meth. Enzymol.* **182**, 428 (1990).
39. J. F. Robyt and B. J. White, *Biochemical Techniques: Theory and Practice*, Waveland Press, Inc., Prospect Heights, Ill., 1987.
40. S. A. Westwood, *Electrophoresis* **6**, (1985).
41. B. Budlowe and R. S. Murch, *Electrophoresis* **6**, 523 (1985).
42. V. Ingram, *Biochim. Biophys. Acta* **28**, 539 (1958).
43. A. Katz, W. Dreyer, and C. Anfinsen, *J. Biol. Chem.* **234**, 2897 (1959).
44. J. W. Jorgenson and K. D. Lukacs, *Anal. Chem.* **53**, 1298–1302 (1981).
45. J. W. Jorgenson and K. D. Lukacs, *Science* **222**, 266–272 (1983).
46. M. J. Gordon, X. Huang, S. L. Pentoney Jr., and R. N. Zare, *Science* **242**, 224 (1988).
47. X. Huang, T.-K. J. Pang, M. J. Gordon, and R. N. Zare, *Anal. Chem.* **59**, 2737–2749 (1987).
48. Y.-F. Cheng and N. J. Dovichi, *Science* **242**, 562 (1988).
49. V. Pretorius, B. J. Hopkins, and J. Schieke, *J. Chromatog.* **99**, 23 (1974).
50. T. Tsuda, K. Nomura, and G. Nakagawa, *J. Chromatog.* **264**, 385 (1983).
51. M. Bier, in J. A. Asenjo and J. Hong, eds., *Separation, Recovery, and Purification in Biotechnology*, ACS Symposium Series No. 314, American Chemical Society, Washington, D.C., 1986, p. 185.
52. W. A. Gobie, J. B. Beckwith, and C. G. Ivory, *Biotech. Prog.* **1**(1), 60 (1985).
53. W. A. Gobie and C. F. Ivory, *AIChE. J.* **34**, 474 (1988).
54. I. Syrový and Z. Hodný, *J. Chromatog.* **569**, 175 (1991).
55. T. Rabilloud, *Electrophoresis* **11**, 785 (1990).
56. C. R. Merrill, M. G. Harasewych, and M. G. Harrington, in M. J. Dunn in Ref. 13, p. 323.
57. H. Ouchterlony, *Acta Pathol. Microbio. Scand.* **25**, 186 (1948).
58. G. Mancini, J. P. Vaerman, A. O. Carbonara, and J. F. Heremans, *Protides Biol. Fluids* **11**, 370 (1963).
59. C.-B. Laurell, *Anal. Biochem.* **15**, 45 (1966).
60. C.-B. Laurell, *Anal. Biochem.* **15**, 45 (1966).
61. Z. L. Awdeh and C. A. Alper, *Proc. Natl. Acad. Sci.* **77**(6), 3576 (1980).
62. J. P. McCue, R. H. Heinard, and R. Tenold, *Rev. Infect. Dis.* **8**(S4), S374 (1986).
63. R. J. Ziccardi, *Clinical Laboratory Techniques for the 1980's*, Alan R. Liss, Inc., New York, 1980, p. 433.
64. C. A. Alper and R. F. Ritchie, in R. F. Ritchie, ed., *Automated Immunoanalysis*, Marcel Dekker, Inc., New York, 1978, p. 139.
65. C. A. Alper and A. M. Johnson, *Vox Sang.* **17**, 445 (1969).
66. P. H. O'Farrell, *J. Biol. Chem.* **250**(10), 4007 (1975).
67. J. E. Celis, B. Honore, G. Bauw, and J. Vandekerckhove, *BioEssays* **12**(2), 93 (1990).

68. E. Southern, *J. Mol. Biol.* **98**, 503 (1975).
69. D. D. Dykes and H. F. Polesky, *Electrophoresis* **6**, 521 (1985).
70. B. Bountin, S. H. Feng, and P. Arnaud, *Am. J. Hum. Genet.* **37**, 1098 (1985).

General References

R. C. Allen, C. A. Saravis, and H. R. Maurer, *Gel Electrophoresis and Isoelectric Focusing of Proteins*, Walter de Gruyter, New York, 1984.
A. Chrambach, *The Practice of Qualitative Gel Electrophoresis*, VCH Publishers, New York, 1985.
B. A. Baldo and E. R. Tovey, *Protein Blotting: Methodology, Research, and Diagnostic Applications*, Karger, Switzerland, 1989.
J. F. Robyt and B. J. White, *Biochemical Techniques: Theory and Practice*, Waveland Press, Inc., Prospect Heights, Ill., 1987.
M. J. Dunn, ed., *Gel Electrophoresis of Proteins*, Wright, Bristol, UK, 1986.
N. Catsimpoolas, *Methods of Protein Separation*, Plenum Press, New York, 1975.
R. F. Boyer, *Modern Experimental Biochemistry*, The Benjamin Cummings Publishing Co., Inc., 1986.
R. J. Hunter, in R. H. Ottewill and R. L. Rowell, eds., *Zeta Potential in Colloid Science, Principles and Applications*, Academic Press, Inc., New York, 1981.

Scott Rudge
Kathleen Markey
Synergen, Inc.

ELECTROSTATIC PRECIPITATION.

See AIR POLLUTION CONTROL METHODS.

ELECTROSTATIC SEALING.

See GLASS.

EMBEDDING

Advances in electronic technology have had great technological and economic impact on the electronic industry throughout the world. The rapid growth of the number of components per chip, the rapid decrease of device dimension, and the steady increase in integrated circuit (IC) chip size (Fig. 1) have imposed stringent requirements, not only on IC physical design and fabrication, but also in electronic packaging and embedding. Electronic embedding is one of the most common processes used to encapsulate and protect these electronic components. With advances in very large-scale integration (VLSI) technology and multichip modules packaging, embedding of this high density package in electronics has become a challenge (1–8). Various polymeric encapsulants are used for embedding these types of electronic components. These materials and their processes are discussed here in detail.

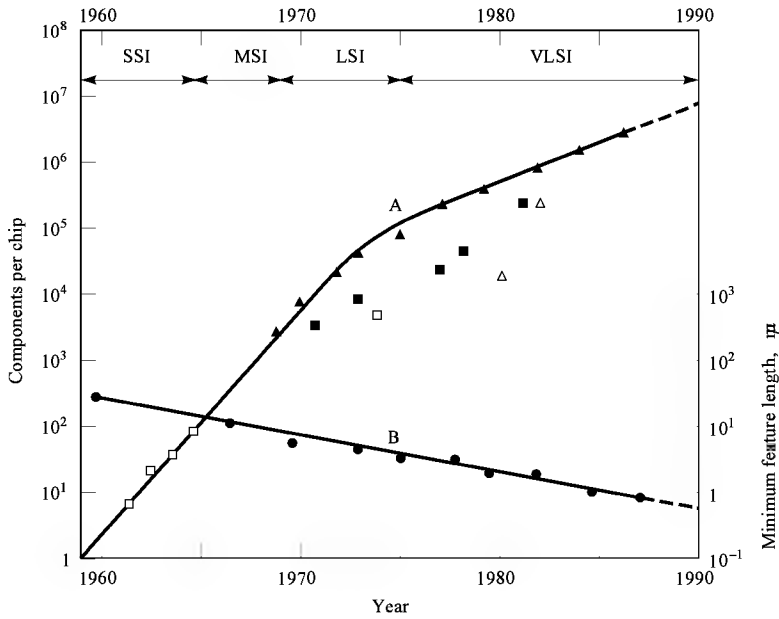


Fig. 1. Technological trends: A, components per chip; B, minimum feature length; ▲, metal oxide semiconductor (MOS) memory; △, bipolar memory; ■, MOS logic; □, bipolar logic. The designations SSI, MSI, LSI, and VLSI stand for small-, medium-, large-, and very large-scale integration, respectively.

Reasons for Electronic Embedding

The purpose of embedding is to protect the electronic components from adverse environments, from ~ 65 to 150°C for military spec 883, thermal shock, temperature cycle during actual life applications, etc; improve handling in later assembly processes; protect line handling and impurities; and increase the long-term reliability of the electronics with a reasonable cost. High performance and low cost materials are the principal driving forces in the 1990s.

Electrooxidation (corrosion) and metal migration are attributed to the presence of moisture. When enough moisture diffuses through the encapsulant to form a continuous water path, with the presence of mobile ion(s) and under an electrical bias, electrocorrosion begins. Figure 2 shows the permeability of various materials. Pure crystals and metals are the best moisture barriers. Glass (silicon dioxide) is an excellent moisture barrier, but it is slightly inferior to pure crystals and metals. Organic polymers, such as fluorocarbons, epoxies, and silicones are a few orders of magnitude more permeable to moisture than glass. Silicone materials have the highest moisture permeability, yet they are one of the best device encapsulants. Obviously, gases are the most permeable to moisture of all materials, as shown in Figure 2. For each particular material the moisture diffusion rate is proportional to the water-vapor partial pressure and inversely proportional to the material thickness. This is accurate when moisture diffusion rates are in steady-state permeation. However, moisture transient penetration rates, perhaps more important because they determine the time it takes for moisture to break through, are inversely proportional to the square of material thickness.

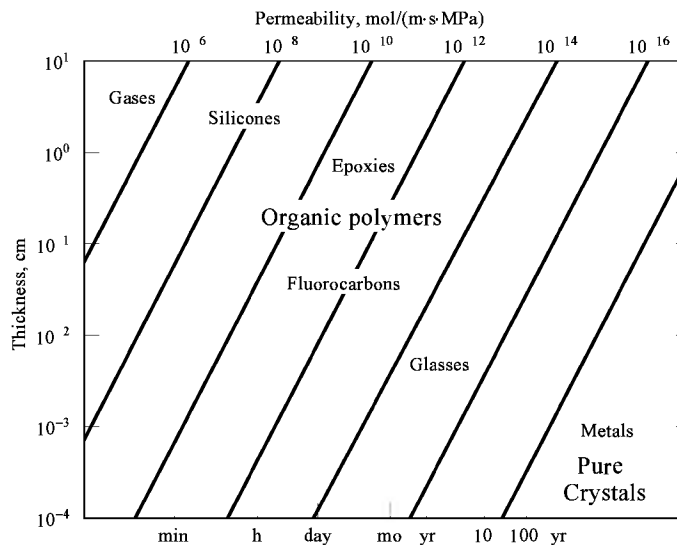


Fig. 2. Moisture permeability of various materials.

Mobile ions, such as sodium or potassium, tend to migrate to the *p-n* junction of the IC device where they acquire an electron, and deposit as the corresponding metal on the *p-n* junction; this consequently destroys the device. Furthermore, mobile ions also support leakage currents between biased device features, which degrade device performance and ultimately destroy the devices by electrochemical processes such as metal conductor dissolution. For example, chloride and fluoride ions, even in trace amounts (ppm), could cause the dissolution of aluminum metallization of complimentary metal oxide semiconductor (CMOS) devices. CMOS is likely to be the trend of VLSI technology and sodium chloride is a common contaminant. The protection of these devices from the effects of these mobile ions is an absolute requirement. The use of an ultrahigh purity encapsulant to encapsulate the passivated IC is the answer to some mobile ion contaminant problems.

Opaque encapsulants can protect light-sensitive optoelectronic devices from uv-vis light radiation damage. However, impurities in an encapsulant, such as traces of uranium in the ceramic or plastic package can cause appreciable alpha particle radiation, as can cosmic radiation in the atmosphere. Alpha radiation can generate a temporary "soft error" in operating dynamic random access memory (DRAM) devices and has become a significant concern, especially in high density memory devices. Good encapsulants must have alpha radiation levels less than 0.001 alpha particles cm² h, and be opaque to protect from uv-vis radiation. Since alpha particles are weakly penetrating, a few micrometers thickness of encapsulant usually prevents this radiation damage to DRAM devices.

Hostile environments, such as extreme cycling temperature (values from -65 to +150 °C in military 883 specs), high relative humidity (85 to 100% o), shock and vibration, and high temperature operating bias are part of real life operation, and the device must survive these operation-life cycles. In addition, encapsulants must also have suitable mechanical, electrical, and physical properties, such as minimal stress and matching thermal expansion coefficient, etc, which are compatible with the IC devices. The encapsulant must have a low dielectric constant to reduce the device propagation delay, and excellent thermal conductivity to dissipate power-hungry, high speed bipolar IC and high density packages which generate tremendous amounts of heat that require special thermal management considerations. Furthermore, since the encapsulation is the final process step and some of the devices are expensive, particularly in high density multichip modules (MCM), it must be easy to apply and repair in production and service. With the proper choice of encapsulant and process, the embedding enhances the reliability of the fragile IC device, and improves its mechanical and physical properties and its manufacturing yields. These are the ultimate goals of the embedding (9-28).

Materials for Electronic Embedding

There are many polymeric materials with suitable electrical, chemical, and physical properties that could be used as electronic components embedding. However, only a few are widely used today.

Silicones. Polydimethylsiloxanes, polydiphenylsiloxanes, and polymethylphenylsiloxanes are generally called silicones (see SILICON COMPOUNDS, SILICONES). With a repeating unit of alternating silicon-oxygen, the siloxane chemical backbone structure, silicone possesses excellent thermal stability and flexibility that are superior to most other materials. Polydimethylsiloxane provides a very low glass-transition temperature *T_g* material but is suitable for use at temperatures up to 200°C. The basis of commercial production of silicones is that chlorosilanes are readily hydrolyzed to give disilanol which are unstable and condense to form siloxane oligomers and polymers. Depending on the reaction conditions, a mixture of linear polymers and cyclic oligomers is produced. The cyclic components can be ring-opened by either acid or base to become linear polymers and it is these linear polymers that are of commercial importance. The linear polymers are typically liquids of low viscosity and, as such, are not suited for use as encapsulants. These must be cross-linked (or vulcanized) in order to increase the molecular weight sufficiently to achieve useful properties. Three methods of cross-linking are used: condensation cures, addition cure systems, and peroxide free-radical cure systems. For electronic applications, only the high purity room temperature vulcanized (RTV) condensation cure silicone, which uses an alkoxide-cure system with noncorrosive alcohol by-products, and platinum-catalyzed addition heat-cure (hydrosilation) silicone systems are suitable for device encapsulation.

The moisture-initiated catalyst, such as organotinates, tin dibutyl dilaurate, etc, assisted RTV process generates water or alcohol by-products which could cause outgassing and voids (Fig. 3). However, by careful control of the curing process, a very reliable encapsulant can be achieved (5-8,29,30). Since the silicone has a low surface tension, it tends to creep and run over the encapsulated IC circuits. To better control the rheological properties of the material, thixotropic agents, such as fumed silica, are usually added to the formulation. The thixotropic agent provides a higher yield stress material, increases the suitable *G'* (storage modulus), *G''* (loss modulus), and *η** (dynamic viscosity) of the encapsulant. Filler-resin and filler-filler interactions are important in obtaining a well-balanced and well-controlled encapsulant. This rheological controlled material tends to flow evenly in each circuit edge, covers all the underchip area, and prevents wicking and runover of the circuits which is a critical parameter in coating production. In addition, pigments such as carbon black and titanium dioxide are usually added as opacifiers to protect light-sensitive devices. Organic solvents such as xylenes and Freons are incorporated into the formulation to reduce the encapsulant viscosity. That is probably one of the reasons why RTV silicone is capable of achieving superior performance in temperature-humidity bias (THB) accelerated electrical tests. Furthermore, the heat curable silicon-hydride and silicon-vinyl addition heat-cure system provides a fast and deep section-cure material system which is preferred in embedding of electronic components.

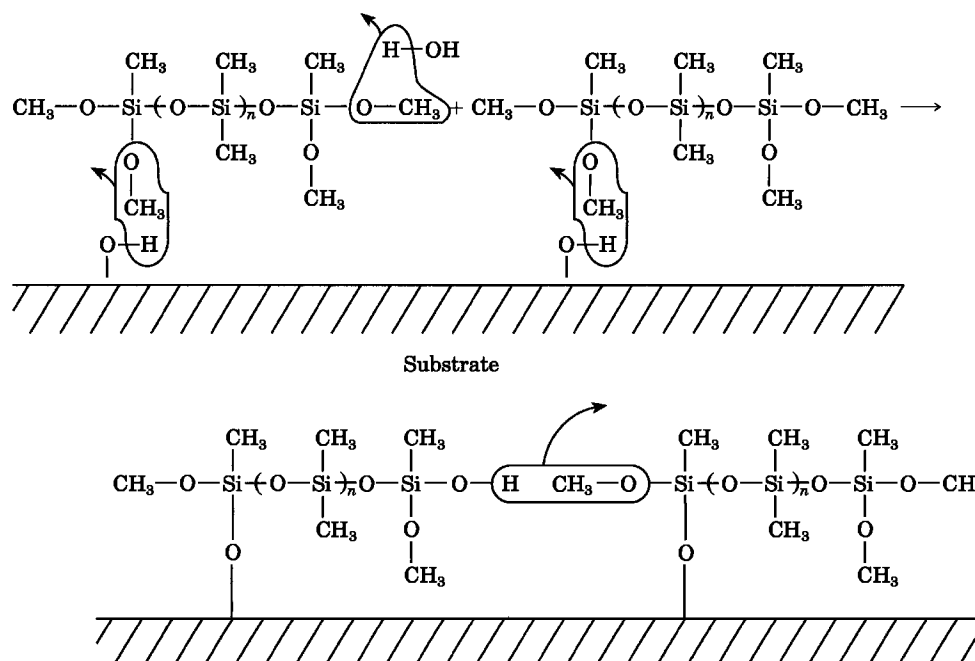


Fig. 3. Condensation-cure mechanism.

Heat-curable hydrosilation silicone (either elastomer or gel) has become an attractive device encapsulant (Fig. 4). Its curing time is much shorter than the RTV-type silicone. With its jelly-like (very low modulus) intrinsic softness, silicone gel is an attractive encapsulant in wire-bonded large-chip-size IC devices. The two-part heat-curable system which consists of the vinyl and hydride reactive functional groups, and the platinum catalyst hydrosilation addition-cure system provide fast-cure systems without any by-product. The key to formulating a low modulus silicone is the deliberate undercross-linking of the silicone system. A few parts per million (ppm) platinum catalyst such as chloroplatinic acid or organoplatinum is used in this system. This catalyst is usually incorporated in the vinyl portion of the resin. This solventless type of heat-curable silicone gel has an increasing use in electronic embedding. The ability of silicone gels to dampen shock and thermal stress is one of the reasons for their use in dedicated wire bonded or flip-chip bonded electronic components (12). Furthermore, the gels are also useful where failure or circuit analysis is needed. The gel allows a probe to be inserted to a test point and then removed; it "repairs" and reseals itself.

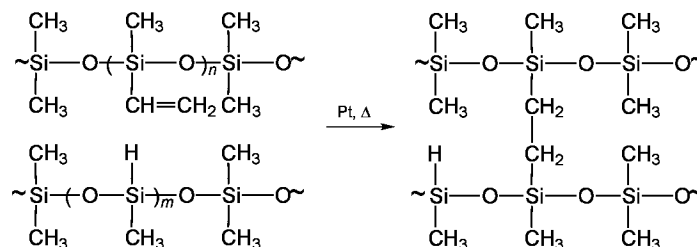


Fig. 4. Addition-cure mechanism, employing hydrosilation involving a vinyl group and hydride; $m > n$.

Silicon-Carbon Thermoset. The Sycar resins of Hercules are silicon-carbon thermosets cured through the hydrosilation of silicon hydride and silicon vinyl groups with a trace amount of platinum catalyst. The material is a fast-cure system (<15 min at 180°C) and shows low moisture absorption that outperforms conventional thermosets such as polyimides and epoxies. Furthermore, the Sycar material provides excellent mechanical and physical properties used in printed wiring board (PWB) laminates and encapsulants such as flow coatable or glob-top coating of chip-on-board type applications.

Epoxies. The unique chemical and physical properties such as excellent chemical and corrosion resistances, electrical and physical properties, excellent adhesion, thermal insulation, low shrinkage, and reasonable material cost have made epoxy resins (qv) very attractive in electronic applications. The commercial preparation of epoxies is based on bisphenol A, which upon reaction with epichlorohydrin produces diglycidyl ethers. The reactant ratio (bisphenol A:epichlorohydrin) determines the final viscosity of the epoxies. In addition to the bisphenol A resins, the novolak resins with multifunctional groups which lead to higher cross-link density and better thermal and chemical resistance have gained increasing acceptance in electronic applications. Typical epoxy curing agents are amines, anhydrides, dicyanodiamides, melamine-formaldehydes, urea-formaldehydes, phenol-formaldehydes, and catalytic curing agents. Anhydrides and amines are two of the most frequently used curing agents.

In most phenoxy resins, the phenolic group is converted into an ether to give improved base resistance. They are cured through the secondary hydroxyl group on the epoxy backbone. High temperature curing is required in this system and it provides excellent chemical resistance. Recently developed high purity epoxies contain greatly reduced amounts of chloride and other mobile ions, such as sodium and potassium, and have become widely used in device encapsulation and molding compounds. The incorporation of large loading (>70% by wt) well-controlled spherical silica particles of silicon dioxide with narrow size-distribution as filler in the epoxy systems has drastically reduced the thermal coefficient of expansion of these materials and makes them more compatible with the IC die-attached substrate materials. The incorporation of a small amount of an elastomeric material, such as silicone elastomeric, carboxyl-terminated butadiene (CTBN) type domain particles, to the rigid epoxy has drastically reduced the elastic modulus, reduced the thermal stress, and increased the toughness of the epoxy material. This new type of low stress epoxy encapsulant has great potential application in molding large IC devices when the epoxy materials are properly formulated and applied, and the stress related issues such as reduced stress and reduced thermal coefficient of expansion have been properly considered and resolved.

Polyurethanes. Polyurethanes were originally based on diisocyanates and diols or polyols (see URETHANE POLYMERS). However, more recent work has focused on the use of intermediates which are low molecular weight polyethers with reactive functional groups such as hydroxyl or isocyanate groups able to further cross-link, chain extend, or branch with other chain extenders to become higher molecular weight polyurethanes. Diamine and diol are chain extended with the prepolymer, either polyester or polyether, to form polyurethanes with urea or urethane linkages, respectively. The morphology of polyurethane is well characterized. Hard and soft segments from diisocyanates and polyols, respectively, are the key to the excellent physical properties of this material (Fig. 5).

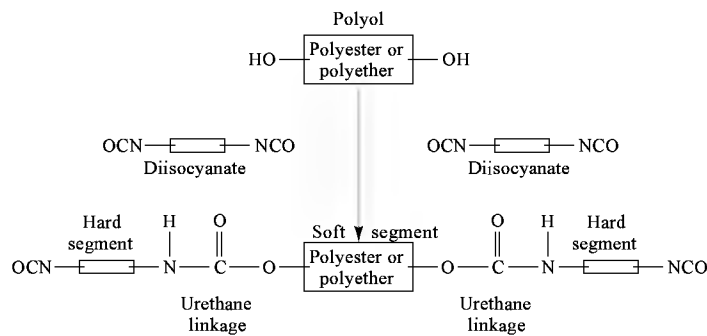
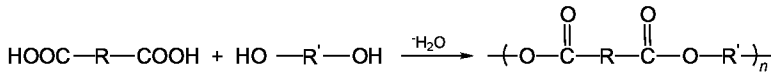


Fig. 5. Synthesis of polyurethane elastomer.

Bases are more widely used than acids as catalysts for polyurethane polymerization. The catalytic activity is increased with the basicity. Amines, such as tertiary alkylamines, and organic metal salts, such as tin or lead octoates, promote the reaction of isocyanate and hydroxyl functional groups in the polyurethane system and accelerate the cross-linking. However, the hydrolytic stability of the polyurethane can be affected by the catalyst and particularly the polyester polyol-based urethanes used. Uv stabilizers are usually added to reduce the radiation sensitivity of the material, especially for those products based on polyols with polybutadiene backbones which do well in hot, humid climates. In addition, polyurethane has unique high strength, high modulus, high hardness, and high elongation. It is one of the toughest elastomers used.

High performance polyurethane elastomers are used in conformal coating, potting, and in reactive injection molding (or reaction impingement molding) of IC devices. Furthermore, rigid polyurethane foams, most often in free-foam densities of 128–288 kg m⁻³ [8–18 lbs ft⁻³ (pcf)], are useful for embedding complex electronic systems. The encapsulant is injected or poured as a liquid into the cavity containing the electronics. As the polymerization reactions occur, a side-reaction with water produces small bubbles and causes the urethane to foam. Usually within two minutes, the rising foam is locked into position by the cure advancement. With proper design, the rising foam pushes air out of the cavity, giving a void-free encapsulant. These materials offer weight reduction, relatively good dielectric properties, and protection from shock. They are used in a number of military ordnance systems at up to 27,000 g-forces.

Polyesters. Polyester is used in embedding resins for electronic components because of its low cost compared to silicones and epoxides. Polyesters (qv) are condensation products of dicarboxylic acids and dihydroxy alcohols; the reaction provides a wide range of viscosities for polyesters.

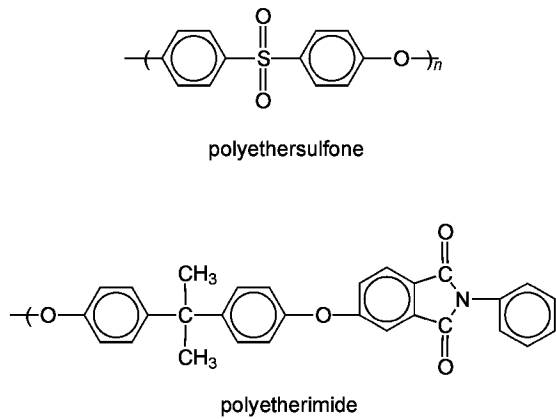


Polyesters may be saturated or unsaturated. The unsaturated polyesters contain double bonds (ethylenic groups) within the polymer backbone structure. These polyesters are cured by cross-linking the unsaturated ethylenic groups with vinyl monomers such as vinyltoluene, divinylbenzene, styrene, and dialkyl phthalate–methyl methacrylate. The reaction mechanism consists of free-radical addition across the double bond with no volatile by-products. Organic peroxides, such as dicumylperoxide and benzoperoxide, are used to promote the free-radical polymerization. Saturated polyesters are produced by the condensation of saturated dicarboxylic acids with polyhydric alcohol. Water is a by-product of the condensation. Their higher shrinkage after cure, by-product water, and high ionic impurity of the materials make polyesters inferior to silicones and epoxies and they cannot be used in critical microelectronic applications. There are electronic-grade polyesters with relatively low viscosities that are suitable for embedding of coils, as in transformers. Furthermore, free-radical-cured polyesters are popularly used for display castings.

Polysulfides. Polysulfides are organic polymers that contain sulfur in disulfide linkages (S–S), mercaptans (S–H), or thiol groups (S–R) (see POLYMERS CONTAINING SULFUR). Low molecular weight polysulfides (3000–4000) may oxidatively react with free-radical reactants such as lead, tin, cobalt octoate, *p*-quinone, dicumene hydroperoxide, lead and manganese dioxides to yield rubbery flexible polysulfides with good water, gas, and moisture sealer properties. Most polysulfides have excellent adhesion to coated substrates and resistance to oxidation, solvents, and ozone. However, the electrical properties of this material are marginal as an insulator, and the dielectric constant of the material is relatively high ($\epsilon_1 < 7$). It is a low cost embedding material for transformers and connector sealing, but not too common for microelectronic embedding.

Advanced Thermoplastics Materials. Thermoplastics and linear plastics of finite molecular weight that can be fabricated into very complex structures by hot melt or injection molding are different from the thermoset materials that require cross-linking to build up infinite molecular weight to form network (cross-link) structures. Advances in thermoplastic engineering materials include amorphous thermoplastics, crystalline thermoplastics, liquid crystal thermoplastics, and fluorinated thermoplastics (see ENGINEERING PLASTICS).

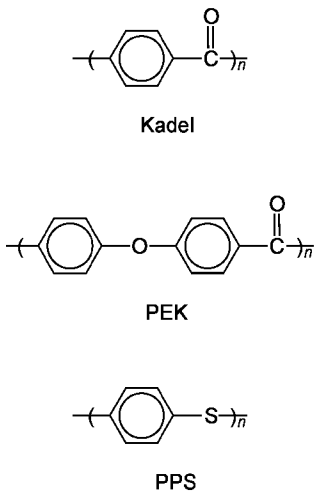
Amorphous Thermoplastics. Polysulfone, polyethersulfone, and polyarylsulfone are examples of amorphous thermoplastics. These materials have high T_g s and can stand temperatures up to 200°C for a long period of time. Amoco and ICI are principal suppliers of this class of material. Imide-based polymers are another class of amorphous thermoplastics.



These materials also have high thermal and oxidative stability. Flexible segments such as amide siloxane can be incorporated into the imide-based structure for hot melt or injection applications. General Electric (GE) and Hoechst-Celanese are suppliers of these high performance plastics.

Crystalline Thermoplastics. These materials are actually semicrystalline, consisting of both crystalline and amorphous regions. The crystalline regions of the material provide a melting point, T_m , in the material, and the amorphous region results in a glass-transition temperature, T_g . In general, T_g is lower than T_m . Unlike amorphous plastics, crystalline plastics experience less property change at T_g . This enhances physical properties between T_g and T_m . The only drawback of this type of plastic material is the high processing temperature required for processing, usually $>T_m$. Two types of

crystalline materials are usually used: ketone-based or poly(phenylene sulfide) (PPS). Polyketone, such as Kadel (Amoco); poly(etherketone) (PEK) such as Hostatec (Hoechst), UltraPek (BASF), and VictrexPek (ICI); poly(etheretherketone) (PEEK) such as Victrex (ICI); and poly(etherketoneketone) such as PEKK (Du Pont) are some of the new generation of ketone-based engineering plastics available for electronic embedding. The most common PPS is Ryton, manufactured by Phillips. Others are Tedur (Bayer), Supec (GE), and Fortron (Hoechst-Celanese). These materials are high performance engineering plastics, with high thermal stability, excellent chemical resistance, and superior mechanical strength.



Liquid Crystal Thermoplastics. Liquid crystal thermoplastics (LCT) have rigid, rod-like backbone structures that tightly pack to form fibrous-like crystalline chains under shear in the melting process. This highly rigid, rod-like structure requires a very high melting temperature to process (>400°C) and creates problems in embedding electronic applications. However, copolymerization with disordered molecules results in lower melting-point materials.

These LCT materials have very high tensile and flexural strength, and excellent mechanical and chemical resistance properties. Some commercial LCT are Vectra (Hoechst-Celanese) and Xydar (Amoco). Du Pont, ICI, GE, and Dow Chemical are also suppliers. Their application in electronic embedding is still in its infancy because of the high temperature processing requirement. Nevertheless, this class of thermoplastic polymers will play an important role in electronic embedding.

Fluorinated Thermoplastics. Teflon, by Du Pont, a poly(tetrafluoroethylene) (TFE), is the best-known fluorinated thermoplastic. The material is highly crystalline with a melting point of 327°C. This high melting point and its high cost limit its use in electronic embedding. However, the material has an exceptionally low dielectric constant and a dissipation factor which is ideal for microwave, high frequency, and superfast clock-rate-device-type applications. Poly(vinylidene difluoride) is another fluorinated thermoplastic that finds use in electronic embedding. Its unique piezo- and pyroelectric behavior makes it an ideal material for sensor transducer applications such as pressure, vibration, infrared, stress-strain, and impact sensing (see FLUORINE COMPOUNDS, ORGANIC POLYTETRAFLUOROETHYLENE and --POLY(VINYLIDENE FLUORIDE)).

Embedding Materials Properties

The ability of a given material to perform as an electronic embedding encapsulant depends largely on its properties. Ultrapure chemical properties with a low level of mobile ions such as sodium, potassium, and chloride are essential. Furthermore, the material's electrical, mechanical, and rheological properties are critical.

Electrical Properties. Electrical properties are important for the corrosion protection of chip-on-board (COB) encapsulated devices. Accelerated temperature, humidity, and bias (THB) are usually used to test the embedding materials. Conventional accelerating testing is done at 85°C, 85% relative humidity, and d-c bias voltage. Triple-track test devices with tantalum nitride (Ta₂N), titanium-palladium-gold (Ti-Pd-Au) metallizations with 76 μm (3 mil) line and spacing are usually used in hybrid integrated circuit (IC) encapsulant testing; however, for IC testing, 5 μm line and spacing metallization test vehicles are generally used. The triple-track test structure is usually biased at the two outer tracks, and the inner track is grounded during the 85°C, 85% rh testing conditions. The leakage current between the outer and inner tracks is monitored periodically with testing time. Good encapsulant, with leakage current < 10⁻⁹ A after 1000 h testing, tends to have a lower leakage current with testing time. The explanation of this phenomena might be attributed to the sweeping out of the contaminant ions during testing (Fig. 6). Furthermore, the change of the resistance of the biased track with respect to the initial resistance of the biased track is also monitored during testing. The increase of the resistances is a direct measurement of the electrooxidation of the biased metallized track. Good encapsulant, with ΔR/R < 1% after 1000 h testing, tends to have lower resistance change with time. In general, high temperature operating bias (HTOB) is also used to weed out the early failure of the imperfect encapsulated devices. Encapsulant plays a critical role in this type of successful electrical testing.

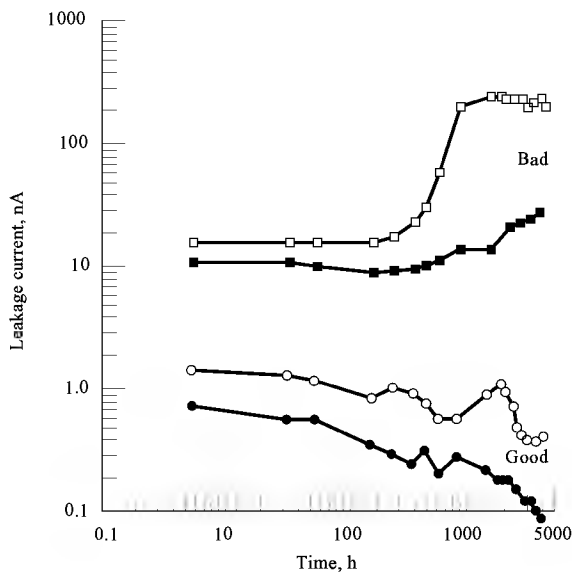


Fig. 6. Temperature-humidity-bias leakage testing of encapsulants at 85°C, 85% rh, and 180 V bias.

Mechanical Properties. Most of electronic IC devices are very fragile. They need strong mechanical protection from the encapsulant to retain their long-term reliability. Encapsulant must provide mechanical protection but still maintain good temperature-cycle and thermal-shock testing, which are part of the routine reliability testing of the embedding electronics.

Temperature-Cycle Testing. Normal high performance electronics require some sort of temperature-cycle testing prior to shipment. The cycling condition varies depending on the degree of reliability. In general, it varies from -30 to 130°C. However, for military applications, 1000 cycles at -65 to 150°C, are required. Temperature-cycle testing is done to weed out those imperfect interconnections made during IC fabrication. Embedding material must be tough and perform well in temperature-cycle testing. Low stress materials are the materials of choice.

Stress of Materials. Low stress materials are essential for dedicated device packaging. To achieve this property a low modulus material with a thermal coefficient of expansion (TCE) match is required. The stress is mainly a result of the following equation:

$$\text{stress} = k \int_{T_1}^{T_2} (\Delta TCE) E dT$$

where *k* is a constant, ΔTCE is the difference in thermal coefficient of expansion, *E* is the modulus of elasticity, and *dT* is the difference in temperature exclusion.

To achieve low stress embedding material, low modulus material such as silicones (elastomers or gels) and polyurethanes are usually used. Soft-domain elastomeric particles are usually incorporated into the hard (high modulus) materials such as epoxies and polyimides to reduce the stress of embedding materials. With the addition of the perfect particle size, distribution, and loading of soft domain particles, low stress epoxy molding compounds have been developed as excellent embedding materials for electronic applications.

Rheological Properties. Materials must have suitable flow properties in order to be used in production. Both Newtonian and non-Newtonian fluids and their viscoelasticity properties play a critical role in the performance of the embedding materials.

Newtonian and Non-Newtonian Materials. A Newtonian material's viscosity is shear-independent, whereas non-Newtonian materials are shear-dependent (Fig. 7). For most potting materials, a Newtonian material is preferred because the material is required to flow under all electronic components, but not be susceptible to shear. However, when flowable material is used for conformal coating applications, a non-Newtonian material with thixotropy agent added is desired since the material should flow on the electronic substrate but stop at the edge without creeping or runover at the circuitry.

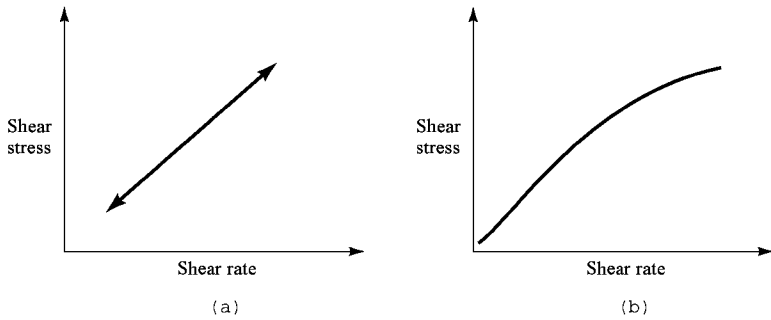


Fig. 7. Viscoelastic behavior of encapsulant materials: (a) Newtonian fluid; (b) non-Newtonian fluid.

Material Processes.

Material processes consist of cavity-filling and saturation coating. The cavity-filling process involves molding, potting, and coating.

Molding. Molding is the most cost-effective and high performance plastic packaging method for IC devices. A polymeric resin (one of the thermosetting molding compounds) is injected into a mold and then cured, ie, the process involves two steps: (1) the molding compound is preheated until it melts and the resin flows through runners, gates, and finally fills up the cavities; and (2) the resin is cured and released from the mold in predetermined shapes. The exact control of the mold pressure, viscosity of the molten molding compound, and the delicate balance of runners, gates, and cavity designs are critical in optimizing the molded plastic IC. Finite element analysis of the plastic molding process is becoming an integral part of improving this process. Since the shear stress of the IC chip molded component could cause wire-bond sweep, device passivation cracks, top-layer metallization deformation, and multilayer oxide and nitride cracks, improved molding compounds and processes could eliminate the damage to the molded IC devices. These molding techniques are well-documented in the literature.

Pressure injection and conformal moldings are some of the current molding processes (Fig. 8). With new advances of low stress molding compounds, techniques such as transfer molding, aperture plate molding, and reactive injection molding are in production use and provide economic ways to encapsulate and package IC devices. A thorough review article and a recently published text offer excellent discussions of molding materials and processes (31,32).

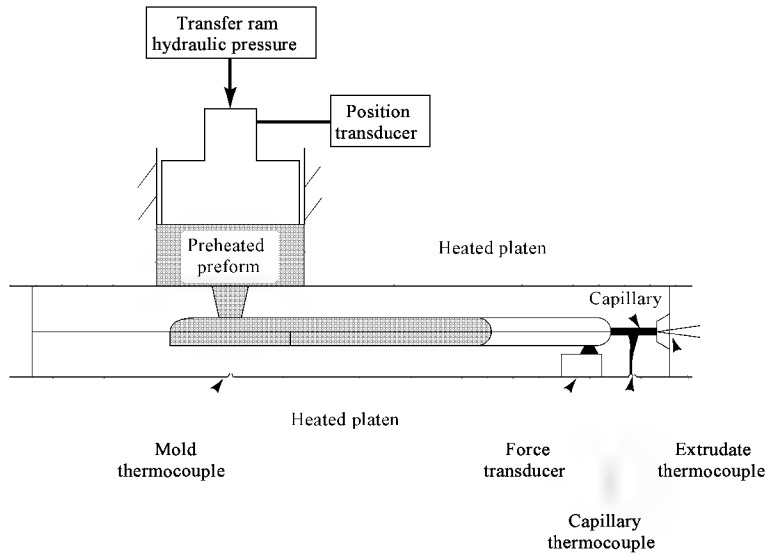


Fig. 8. Typical conformal molding process.

Potting. Potting is the simplest cavity-filling process. It involves placing the electronic component within a container, filling the container with a

liquid resin, and then curing the material as an integral part of the component. Polymeric resins, such as epoxies, silicones, and polyurethanes, are usually used as potting materials. Containers such as metal cans or rugged polymeric casings made from high performance engineering thermoplastic polymers, enhance the effectiveness of the encapsulant. Adhesion between the potted material and the casing is essential in achieving long-lasting reliable packages. In the fast-growing automated manufacturing process, rugged, machine-insertable components, such as surface-mounted chip carriers, dual-in-line (DIP), single-in-line (SIP) packages, molded and potted packages, and discrete components are highly desirable.

Casting. Casting is similar to potting, except the outer case is removed after the polymer cavity-filling process is completed and cured. No heat or pressure is applied in the process. This labor intensive casting process is not commonly used, as compared to potting and molding of modern electronic packages.

Saturation and Coating Processes. Saturation coatings consist of impregnation, dip coating, and conformal and surface coatings.

Impregnation Coating. Impregnation coating is performed by saturation of the component with a low viscosity resin; a thin film is coated on the component surface. This process is usually used with a cavity-filling or conformal coating process.

Dip Coating. Dip coating is performed by dipping the component into an encapsulating resin. The component is then withdrawn, dried, and cured. Coating thickness is usually a function of resin viscosity, withdrawal rate, and coating speed. This process also depends on the resin reactivity, curing rate, and curing temperature, etc. The dip coating process is widely used in glass-laminated printed circuit board and optical fiber coatings, where a fast-cure resin is required for the process.

Conformal and Surface Coatings. Conformal and surface coatings are the common techniques used in IC device encapsulation, and include spin coating, spray coating, and flow coating of the encapsulant onto the component. Suitable rheological properties of the encapsulant such as dynamic viscosity (η), yield stress, G' (storage modulus), and G'' (loss modulus) are critical in obtaining a good flow-coating package (Fig. 9), especially in hybrid IC encapsulation, where the encapsulant tends to runover from the substrate and wick onto the leads of the hybrid devices. In addition, fluidized epoxy powder bed coating of single-in-line package (SIP) hybrid ICs and printed wiring boards (PWBs) are an attractive conformal coating process. Surface-mounted components on PWB are routinely encapsulated by this conformal coating process.

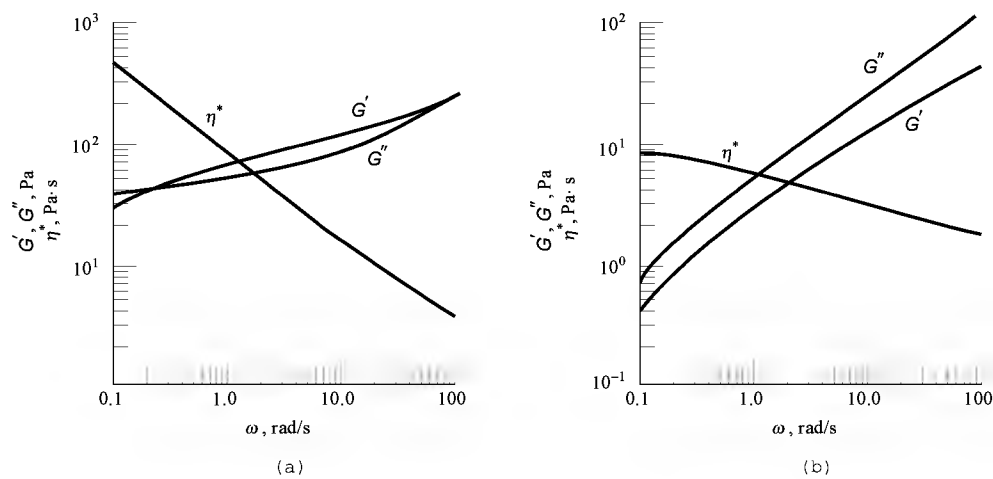


Fig. 9. Rheology of silicone encapsulants: (a) good material; (b) bad material. To convert Pa·s to P, multiply by 10. To convert Pa to dyn/cm², multiply by 10 (22).

Material Curing

In order to optimize each embedding material property, complete cure of the material is essential. Various analytical methods are used to determine the complete cure of each material. Differential scanning calorimetry, Fourier transform-infrared (ftir), and microdielectrometry provide quantitative curing processing of each material. Their methods are described below.

Differential Scanning Calorimetry. A dsc measurement shows the heat flow to or from the material during its curing process. Heat capacity is measured and quantitative information regarding enthalpic transitions in a material is provided (Fig. 10). A dsc thermogram is a plot of energy supplied to or by the material as a function of temperature. The peak areas can be related to the enthalpic changes quantitatively. The T_g (glass-transition temperature) can be taken as the temperature at which one-half change in heat capacity (ΔC_p) (or the temperature at the midpoint of the change between the specific heats of the glassy and rubbery states) has occurred. Various kinetic properties can also be extracted from the dsc, eg, activation energy (E_{act}), reaction rate and reaction order. When materials are fully cured, there should be no discernable exothermic or endothermic peak as the temperature of the sample is raised at a programmed rate. In general, a sample size of a few milligrams is needed for such a dsc measurement.

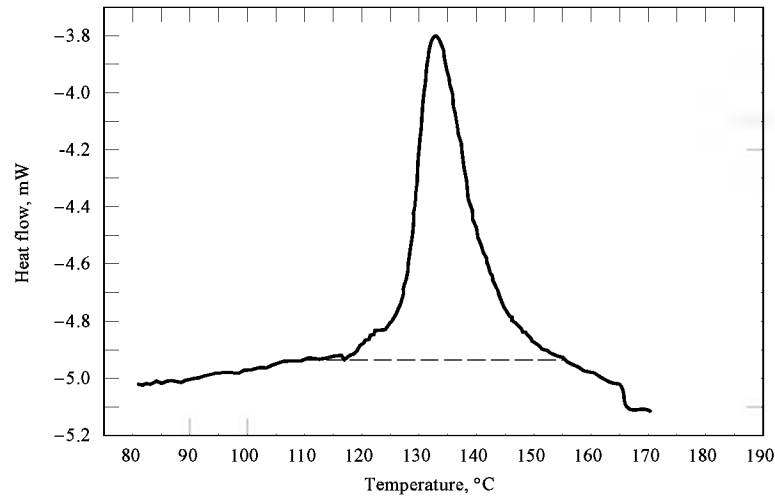


Fig. 10. Differential scanning calorimetry thermogram of a thermoset. The reaction order is 1.83, E_{act} = 95 kJ/mol (22.7 kcal/mol), and the heat of reaction is 6.95 J/g (1.66 cal/g).

Fourier Transform Infrared (ftir) Spectroscopy. Ftir is a sensitive tool for measuring the vibrational energy of the reactive functional

groups of the encapsulant such as in heat curable silicones. For example, strong Si-H absorption at $\sim 2100\text{ cm}^{-1}$ shows the presence of a large amount of the uncured silicone hydride. Upon hydrosilation cure, the decrease of Si-H absorption can be easily measured by its peak height or peak area integration and thus provides a useful tool for quantifying the material cure. Stabilization of the Si-H absorption is a good indication of the complete cure of the material.

Microdielectrometry. Microdielectrometry is another technique for quantitative characterization of encapsulant cure. Micromet Instruments's Eumetric System II microdielectrometer or TA Instruments's Dynamic Mechanical Thermal Analyzer with a miniature IC sensor and a wide range of frequencies (from 0.05 Hz to 10 kHz) may be used to monitor the loss factor (E'') of the encapsulant. This is a sensitive technique to detect the final stage of the material cure. A thin layer of freshly prepared specimen ($\sim 500\text{ }\mu\text{m}$ thickness) is coated on the miniature IC sensor and placed inside a programmable oven. The temperature of the oven is set to the predescribed temperature, ie, 120, 150, or 175°C, and the loss factor (E'') and dielectric constant (E') at various frequencies (0.05, 1, 100, 1,000, 10,000 Hz) are monitored periodically during the curing time (Fig. 11). When all the E' and E'' measurements are stabilized during the isothermal cure, it is a good indication of a complete cure. Low frequency (less than 0.01 Hz) sweep experiments provide a sensitive tool in quantifying the material cure (30).

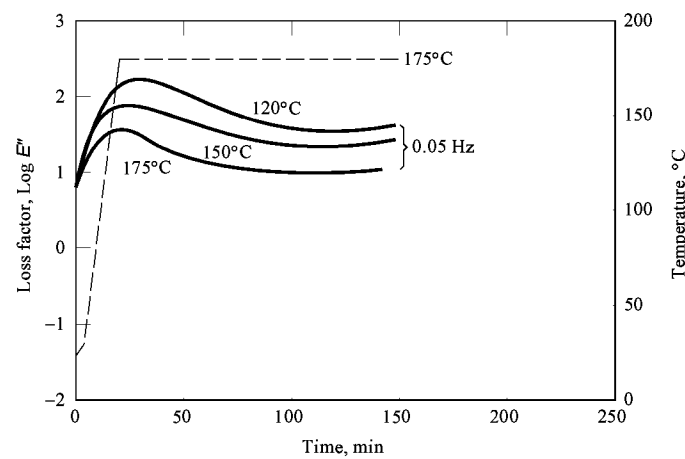


Fig. 11. Microdielectrometry study of a thermoset resin.

BIBLIOGRAPHY

"Embedding" in *ECT* 2nd ed., Vol. 8, pp. 102–116, by F. J. Modic and D. A. Barsness, General Electric Co.; in *ECT* 3rd ed., Vol. 8, pp. 877–899, by C. A. Harper, Westinghouse Electric Corp.

1. C. P. Wong, in C. P. Wong, ed., *Polymers for Electric and Photonic Applications*, Academic Press, San Diego, Calif., 1993, Chapt. 4, pp. 167–214.
2. C. A. Harper, ed., *Electronic Packaging and Interconnect Handbooks*, McGraw Hill Book Co., New York, 1991.
3. M. L. Minges, ed., *Electronic Materials Handbook*, ASM International, Metals Park, Ohio, 1989.
4. R. R. Tummala and E. J. Rymaszewski, eds., *Microelectronic Packaging Handbook*, Van Nostrand Reinhold Co., Inc., New York, 1989.
5. C. P. Wong, in J. H. Lai, ed., *Polymers in Electronics*, CRC Press, Boca Raton, Fla., 1989, Chapt. 3, pp. 63–92.
6. C. P. Wong, *Application of Polymers in Encapsulation of Electronic Parts*, *Advances of Polymer Science*, Vol. 84, Springer-Verlag, Berlin, 1988, pp. 63–83.
7. C. P. Wong in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 5, John Wiley & Sons, Inc., New York, 1986, p. 638.
8. D. Seraphim, R. Ladky, and C-Y. Li, *Principles of Electronic Packaging*, McGraw Hill Book Co., Inc. New York, 1989.
9. M. C. Volk and co-workers, *Electrical Encapsulations*, Van Nostrand Reinhold, New York, 1962.
10. W. Noll, *Chemistry and Technology of Silicone*, Academic Press, Inc., New York, 1968.
11. D. J. Shanefield and W. Colling in *Proceedings of 1st International Society for Advancement of Materials and Processing Engineers (SAMPE) Electronics Conference*, Covina, Calif., 1987, p. 70.
12. C. P. Wong in R. Jacodine, K. A. Jackson, and R. C. Sundahl, eds., *Material Research Society Symposium Proceedings*, Pittsburgh, Pa., Vol. 108, 1988, pp. 175.
13. F. N. Sinnadurai, *Handbook of Microelectronics Packaging and Interconnection Technologies*, Electrochemical Publications Limited, 1985, p. 8.
14. R. Miller in *Proceedings of the National Electronic Packaging and Production Conference (NEPCON) West*, Des Plaines, Ill., 1986, pp. 867–870.
15. C. P. Wong, *Int. J. Hybrids Microelectron.* **4**(2), 315 (1981).
16. C. P. Wong and D. E. Maurer in *Semiconductor Moisture Measurement Technology*, Special Publication 400-72, National Bureau of Standards, 1982, p. 275.
17. C. P. Wong, *Improved Room-Temperature Vulcanized Silicone Elastomers as Integrated Circuit Encapsulants*, *Polymer Materials for Electronics Applications*, American Chemical Society Symposium Series, Washington, D.C., Nos. 184, 171, 1982.
18. H. Danielsson in *International Society for Hybrids and Microelectronics Proceedings*, Reston, Va., 1988, p. 135.
19. M. L. White, *Proc. IEEE* **27**, 1610 (1969); J. H. Martin and L. D. Hanley, *IEEE Trans. Comp. Hybrids Manuf. Technol.* **CHMT-4**, 210 (1981).
20. R. G. Mancke, Ref. 19, 492 (1981).
21. N. Sinnadurai, *Microelectron. J.* **12**, 6 (1981).
22. C. P. Wong and D. M. Rose, *IEEE Trans. Comp. Hybrids Manuf. Technol.* **CHMT-6**, 485–493 (1983).
23. C. P. Wong, in *Proceedings of the IEEE 5th VLSI Packaging Workshop*, Paris, France, 1986, p. 45.
24. R. Miller, *ISHM J.* **4**, 315 (1986).
25. K. Otsuka and co-workers in *Proceedings of the International Electronic Packaging Society*, Des Plaines, Ill., 1986, p. 720.
26. K. Otsuka and co-workers, *IEEE Trans. Comp. Hybrids Manuf. Technol.* **CHMT-12** (1987).
27. C. P. Wong, *Material Res. Soc. Proc.* **108**, 175 (1988).
28. C. P. Wong, J. M. Segelken, and J. W. Balde, *IEEE Trans. Comp. Hybrids Manuf. Technol.* **CHMT-12**(4), 419 (1989).
29. C. P. Wong, *Polymers in Electronics*, American Chemical Society Symposium Series, Washington, D.C., No. 242, 1984, Chapt. 23, p. 285.
30. C. P. Wong, *Polymer Science and Engineering Proceedings*, Vol. 55, American Chemical Society, Washington, D.C., 1986, p. 803.
31. N. Kinjo, M. Ogata, K. Nishi, and A. Kaneda, *Adv. Polym. Sci.* **88**, 1 (1989).
32. L. T. Manzione, *Plastic Packaging of Microelectronic Devices*, Van Nostrand Reinhold, New York, 1990.

C. P. Wong
Paul D'Ambra

AT&T Bell Labs

EMULSIONS

The traditional view of emulsion stability (1,2) was concerned with systems of two isotropic, Newtonian liquids of which one is dispersed in the other in the form of spherical droplets. The stabilization of such a system was achieved by adsorbed amphiphiles, which modify interfacial properties and to some extent the colloidal forces across a thin liquid film, after the hydrodynamic conditions of the latter had been taken into consideration. However, a large number of emulsions, in fact, contain more than two phases. The importance of the third phase was recognized early (3) and the IUPAC definition of an emulsion included a third phase (4). With this relation in mind, this article deals with two-phase emulsions as an introduction. These systems are useful in discussing the details of formation and destabilization, because of their relative simplicity. The subsequent treatment focuses on three-phase emulsions, outlining three special cases. The presence of the third phase is shown in order to monitor the properties of the emulsion in a significant manner.

Simple Emulsions

A (macro)emulsion is formed when two immiscible liquids, usually water and a hydrophobic organic solvent, an oil, are mechanically agitated (5) so that one liquid forms droplets in the other one. A microemulsion, on the other hand, forms spontaneously because of the self-association of added amphiphilic molecules. During the emulsification agitation both liquids form droplets, and with no stabilization, two emulsion layers are formed, one with oil droplets in water (o w) and one of water in oil (w o). However, if not stabilized the droplets separate into two phases when the agitation ceases. If an emulsifier (a stabilizing compound) is added to the two immiscible liquids, one of them becomes continuous and the other one remains in droplet form.

The type of emulsion (o w or w o) is determined by the volume ratio of the two liquids provided the ratio is sufficiently high. For example, with 5% water and 95% oil (an o w phase ratio of 19), the emulsion will become w o unless extreme measures are taken to ensure the formation of an o w emulsion.

For moderate ratios (<3), the type of emulsion is decided by several factors (5), such as order of addition or type of emulsifier. One liquid slowly added to the other with agitation usually results in the last-mentioned phase being the continuous one. Another factor is preferred solubility of the emulsifier; the phase in which the emulsifier is soluble most probably is continuous.

The most common types of emulsions consist of only two liquids, water and an oil. An o w emulsion consists of oil droplets dispersed in a continuous aqueous phase, and a w o emulsion consists of water droplets dispersed in oil (Fig. 1). Occasionally inversion takes place; an o w emulsion changes into w o emulsion and vice versa. More complex emulsions such as double emulsions are formed because the water droplets in a continuous oil phase themselves contain dispersed oil droplets (Fig. 2). Such oil-in-water-in-oil emulsions are noted as o w o. In the same manner a w o w emulsion may be formed, which finds use as a system for slow delivery, extraction, etc (6,7).

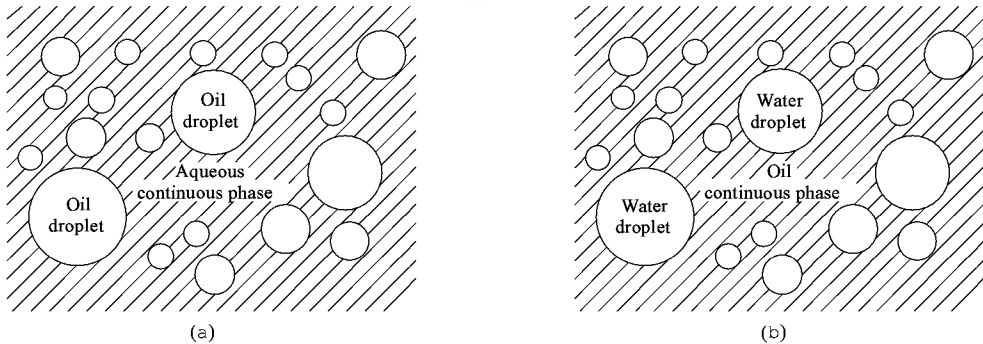


Fig. 1. In an oil-in-water (o w) emulsion (a), macroscopic oil droplets are dispersed in water. In a water-in-oil (w o) emulsion (b), the situation is reversed.

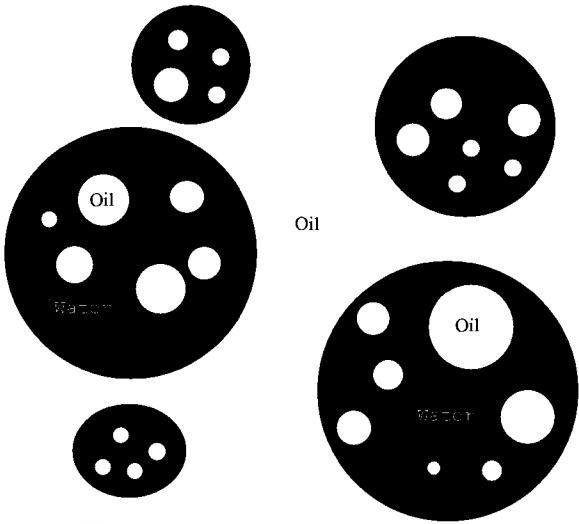


Fig. 2. In an oil-in-water-in-oil (o w o) emulsion the water droplets in the oil phase themselves contain oil droplets.

During emulsification new surfaces are created between the two phases. Such a process requires energy; the surface free energy, numerically identical to the easily measured surface tension, reflects this amount.

There are two fairly common misconceptions about the conclusions that may be drawn from the value of the surface free energy. First, the energy input to enlarge the interface during emulsification is not a significant part of the total energy needed for this process. The viscous resistance during the agitation absorbs most of the energy, giving a small temperature rise to the system. Secondly, the emulsion is stabilized by the addition of a compound adsorbed to the interface, termed an emulsifier. Emulsifiers are molecules with nonpolar and polar parts that reside at the interface. Their presence causes a reduction of the surface tension and a stabilization of the emulsion, but these two features are not directly related. The interfacial tension reflects the amount of emulsifier at the interface.

Any conclusion that a low interfacial tension per se is an indication of enhanced emulsion stability is not reliable. In fact (8), very low interfacial tensions lead to instability. The stability of an emulsion is influenced by the charge at the interface and by the packing of the emulsifier molecules, but the interfacial tension at the levels found in the common emulsion has no influence on stability.

Emulsification

Much commercial equipment is available for emulsification (Fig. 3) and has been well described (9). The following discusses the relations between energy input and emulsion droplet size.

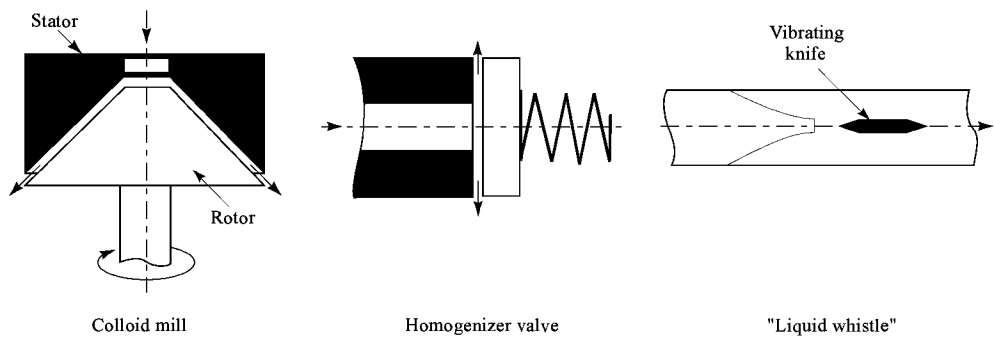


Fig. 3. Commercial emulsification equipment is built on different principles for mixing one liquid into another (11).

The emulsification process in principle consists of the break-up of large droplets into smaller ones due to shear forces (10). The simplest form of shear is experienced in lamellar flow, and the droplet break-up may be visualized according to Figure 4. The phenomenon is governed by two forces, ie, the Laplace pressure, which preserves the droplet, and the stress from the velocity gradient, which causes the deformation. The ratio between the two is called the Weber number, We , where η_c is the viscosity of the continuous phase, G the velocity gradient, r the droplet radius, and γ the interfacial tension.

$$We = \frac{\eta_c \cdot G \cdot r}{\gamma}$$
(1)

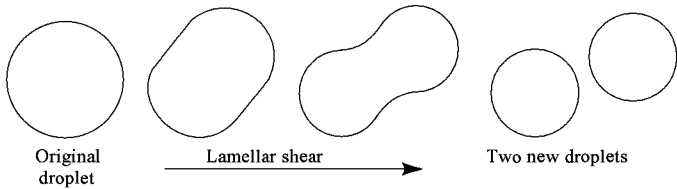


Fig. 4. A droplet is broken into two droplets in lamellar shear.

As an approximate rule, break-up of droplets occurs for a Weber number in excess of one, a rule of thumb that is actually valid for the range of viscosity ratios of the dispersed phase to the continuous phase of less than approximately five. Higher viscosities of the disperse phase lead to serious difficulties with emulsification because the shear energy is then dispersed in rotation of the droplets.

As may be expected, turbulent flow (9,11) is more efficient for droplet formation in low viscosity liquids. With the average amount of energy dissipated per unit time and volume equal to Σ and mass density equal to ρ , the larger eddies are characterized by a velocity gradient equal to

$$u(x) \sim \Sigma^{1/3} X^{2/3} \rho^{-1/3}$$
(2)

The pressure fluctuations caused by the eddies are of the order of $\rho[u(x)]^2$ and if these exceed the Laplace pressure a neighboring droplet of diameter x will be broken up. The result is an upper limit to size.

$$d_{\max} \sim \Sigma^{-2/5} \gamma^{3/5} \rho^{-1/5}$$
(3)

This relation also holds for the average droplet size and a doubling of Σ with other factors retained causes a reduction in average droplet size by 25%. Σ is the energy density per time and this feature is to the advantage of emulsification using a homogenizer as compared to the process with a stirrer (Fig. 3). The former dissipates the same amount of energy in a much shorter time, causing a significant increase in Σ . In addition, the homogenizer has the advantage of an increase in the pressure (power consumption) which additionally shortens the time through the homogenizing valve; eg, Σ is increased. Hence, the average diameter does not depend on the power consumption with an exponent 0.4 but rather 0.6 (10). These features indicate that enhanced duration (batch) or repetitious (continuous) treatment are not economically efficient, an estimation borne out by experimental results (12).

One consequence of the Σ dependence is that the higher energy density per volume may be used to advantage by emulsification of the dispersed phase into a reduced amount of the continuous phase, followed by dilution. A reduced amount of the continuous phase means an increased value of Σ , because the energy input is dissipated into a smaller volume. An exception to this rule is found if the continuous phase contains solid particles. In such a case forces acting through the liquid medium are not efficient for obvious reasons, and mechanical means such as a roller mill should be used.

Finally, some general rules for the amount of surfactant appear to be valid (13). For anionic surfactants the average size of droplets is reduced for an increase of surfactant concentration up to the critical micellization concentration, whereas for nonionic surfactants a reduction occurs also for concentrations in excess of this value. The latter case may reflect the solubility of the nonionic surfactant in both phases, causing a reduction of interfacial tension at higher concentrations, or may reflect the stabilizing action of the micelles per se.

Stability of Two-Phase Emulsions

In a two-phase emulsion, if the material is allowed (or forced) to separate completely, only two transparent layers are found. The emulsifiers in the form of

surfactants or polymers are soluble in the water, the oil, or both and are found in one or both clear phases, when the emulsion is separated.

Before determining the degree of stability of an emulsion and the reason for this stability, the mechanisms of its destabilization should be considered. When an emulsion starts to separate, an oil layer appears on top, and an aqueous layer appears on the bottom. This separation is the final state of the destabilization of the emulsion; the initial two processes are called flocculation and coalescence (Fig. 5). In flocculation, two droplets become attached to each other but are still separated by a thin film of the liquid. When more droplets are added, an aggregate is formed, in which the individual droplets cluster but retain the thin liquid films between them, as in Figure 5a. The emulsifier molecules remain at the surface of the individual droplets during this process, as indicated in Figure 6.

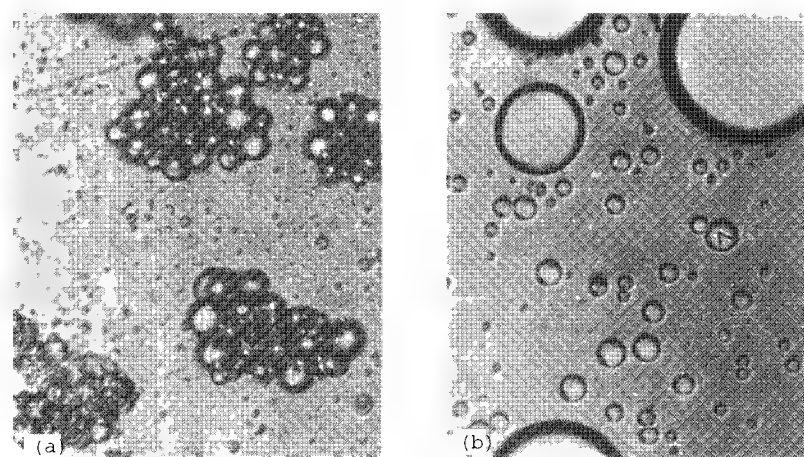


Fig. 5. (a) In a flocculated emulsion the droplets are aggregated but separated by a thin film. Coalescence means that the thin film between droplets bursts and the aggregates become single droplets. (b) The result is an emulsion with a wide distribution of droplet sizes but without aggregates.

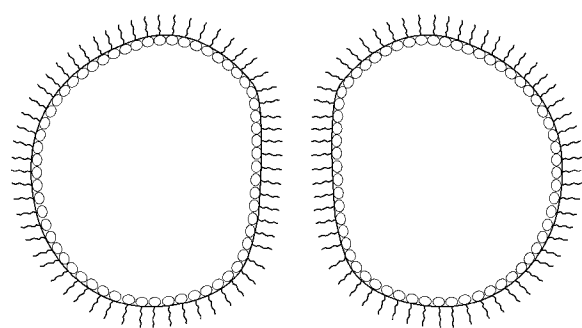


Fig. 6. In the flocculated droplets the emulsifier molecule remains at the oil–water interface.

In the coalescence step, the thin liquid film between the droplets is destabilized, and a large droplet is formed. Hence, the coalescing emulsion is characterized by a wide size distribution of the droplets, but no clusters are present (Fig. 5b). Finally, the droplets achieve such a size that they are recognized by the naked eye as a separate phase. A fully separated emulsion consists of an oil layer and an aqueous layer.

The sequence, flocculation → coalescence → separation, is complicated by the fact that creaming or sedimentation occurs and that this process is determined by the droplet size. The sedimentation velocity is monitored by the oppositely directed forces which form the buoyancy and the viscous drag of the continuous phase on the droplet:

$$f_b = 4\pi r^3 \Delta\rho \cdot g \qquad f_d = 6\pi\eta r v \tag{4}$$

where f_b is the buoyancy force, r the droplet radius, $\Delta\rho$ the difference in density, f_d the drag force, η the viscosity of the continuous phase, and v the sedimentation velocity. v is given by equation 5 and the sedimentation velocity is proportional to the radius squared.

$$v = \frac{2r^2 \Delta\rho \cdot g}{9\eta} \tag{5}$$

There are two essential consequences of this relation. Because larger droplets sediment or rise much faster (a 5- μm drop rises 625 times faster than a 0.2- μm droplet), the process is equal to shearing, leading to enhanced flocculation. The ratio between flocculation due to shear and to diffusion of droplets is proportional to the cube of the radius. Secondly, flocculation to droplet aggregates means an enhanced sedimentation rate. Six drops in an octahedral arrangement gives approximately four times the sedimentation rate.

The definition of stability and the appropriate measure of the rate of destabilization of an emulsion depend on the application. For an application such as the use of a fluorocarbon as a blood substitute the destabilization stage of importance is obviously the flocculation; aggregates of droplets would clog blood vessels. The stabilization must be efficient at long-range distances to prevent the droplets from reaching each other. At the other extreme are beverages such as soda or reconstituted fruit juices in which aggregation of flavor droplets is not noticed by the consumer and is not an essential disadvantage. The consumer instead would notice any separated oil forming a "greasy" ring in the bottle neck. In this case creaming must be prevented, a distinctly different problem from stabilization against flocculation.

STABILIZATION MECHANISMS

The stabilization of an emulsion involves slowing the destabilization, primarily the flocculation process. This may be achieved in two principal manners: by reducing the mobility of droplets through enhanced viscosity or by inserting an energy barrier between them (see also DISPERSANTS; FLOCCULATING AGENTS).

Viscosity Increase. The flocculation rate of an emulsion is inversely proportional to the viscosity of the continuous phase and an increase of the viscosity from 1 mPa·s (cP) (water at room temperature) to a value of 10 Pa·s (100 P) (waxy liquid) reduces the flocculation rate by a factor of 10,000. Such a change would give a half-life of an unprotected emulsion of a few hours, which is of little practical use.

To prepare stable emulsions in this way gelation of the continuous medium is necessary. The appearance of a liquid emulsion may be retained by choosing a polymer for the continuous phase, giving a thixotropic solution with short breakdown and buildup times. The polymers used for this purpose are natural gums (qv) or synthetic polymers. Clay particles also act as viscosity enhancers. The members of the bentonite family derived from

montmorillonites swell in water and strongly enhance the viscosity at pH values in excess of 6 (see CLAYS). The method of an energy barrier between droplets is the more common method to stabilize emulsions.

Energy Barrier. An energy barrier results in repulsive forces between droplets when they approach each other. The half-life of an emulsion changes drastically when energy barriers of different height are introduced in the system. Increasing the barrier to 20 kT gives a typical half-life of a few years which is sufficient for most applications.

$W, \text{ kT}$	$t_{1/2}$
0	$\approx /s$
5	$\approx 40 \text{ s}$
10	$\approx 1.5 \text{ h}$
20	$\approx 4 \text{ yr}$
50	$4 \times 10^{13} \text{ yr}$

Two kinds of barriers are important for two-phase emulsions: the electric double layer and steric repulsion from adsorbed polymers. An ionic surfactant adsorbed at the interface of an oil droplet in water orients the polar group toward the water. The counterions of the surfactant form a diffuse cloud reaching out into the continuous phase, the electric double layer. When the counterions start overlapping at the approach of two droplets, a repulsion force is experienced. The repulsion from the electric double layer is famous because it played a decisive role in the theory for colloidal stability that is called DLVO, after its originators Derjaguin, Landau, Vervy, and Overbeek (14,15). The theory provided substantial progress in the understanding of colloidal stability, and its treatment dominated the colloid science literature for several decades.

DLVO THEORY

The theory has certain practical limitations. It is useful for o w (oil-in-water) emulsions but for w o (water-in-oil) systems DLVO theory must be applied with extreme caution (16). The essential use of the DLVO theory for emulsion technology lies in its ability to relate the stability of an o w emulsion to the salt content of the continuous phase. In brief, the theory says that electric double-layer repulsion will stabilize an emulsion, when the electrolyte concentration in the continuous phase is less than a certain value.

The theory relates the stability of emulsified droplets to two independent potentials that come into action when two droplets approach each other. The van der Waals potential is negative (attractive) and the potential from the electric double layer is positive (repulsive), as shown in Figure 7. The resulting energy is the sum of the two energies. In the example the energy from the electric double layer is numerically greater than the van der Waals potential for a certain distance range, and the sum of the two energies becomes positive, giving a repulsive resulting force. However, for small distances the resulting energy is always negative, and the particles irreversibly aggregate if the distance is reduced to a value less than that corresponding to the maximum of the combined potential. The maximum of the combined potential serves as the energy barrier discussed earlier; a barrier height of 20 kT is sufficient for stability in most particle emulsions systems.

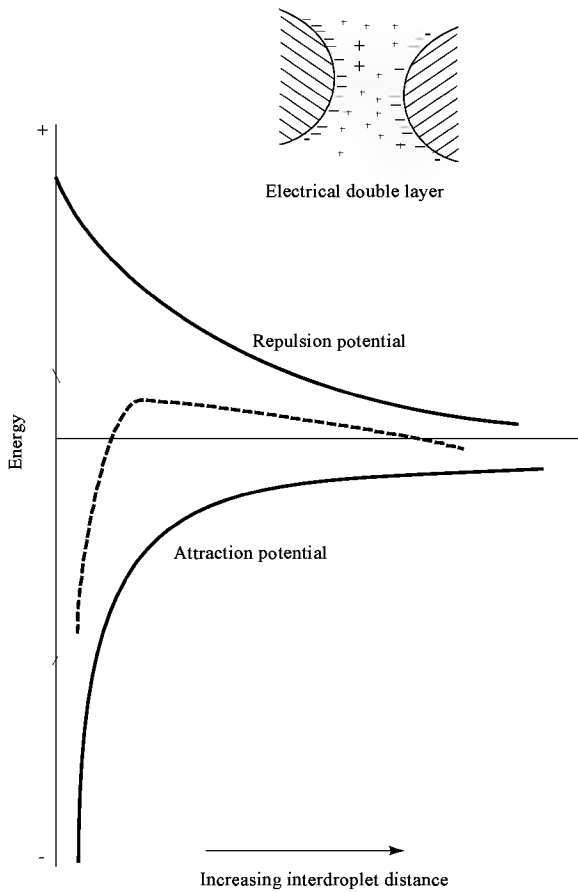


Fig. 7. The van der Waals potential between droplets is increasingly negative with reduced interdroplet distance, whereas the electric double-layer potential is increasingly positive. The resulting potential (---) may have a maximum.

This energy maximum is calculated from the electric surface potential. An approximation of this surface potential is the zeta potential, which is experimentally determined with commercial instruments. For o w emulsions with low electrolyte content in the aqueous phase, a zeta potential of 40 mV is sufficient to bring the energy maximum to this level.

The relative value of the two potentials reveals the destabilization action of salts added to the emulsion. Addition of an electrolyte to the continuous phase causes a reduction of the electric double-layer repulsion potential, whereas the van der Waals potential remains essentially unchanged. Hence, the reduced electric double-layer potential causes a corresponding reduction of the maximum in the total potential, and at a certain concentration of electrolyte the maximum barrier height is reduced to a level at which the stability is lost.

Calculations of the relative height of the barrier with electrolyte content are interesting from a scientific point of view, but of limited value in

formulation efforts. The essential information in practice is that an oil emulsion can be stabilized by the repulsion from the electric double layer only if the aqueous phase contains monovalent counterions at a concentration less than 0.1 M, divalent ones at less than 0.01 M, or trivalent ones at less than 0.001 M. For these cases a zeta potential of 40 mV or greater provides what stability the electric double layer can offer. For electrolyte concentrations higher than these values or for a water emulsion, the zeta potential has no importance for stability; other stabilizing agents must be found.

As a related matter it is easily understood that addition of salts at a certain concentration destabilizes an emulsion. It may be concluded that if an emulsion remains stable at electrolyte contents higher than those cited in the preceding paragraphs, the stability is not the result of electric double-layer repulsion, which may be useful information to find the optimum manner for destabilization.

A special case of destabilization is by hydroxy complex ions. Aluminum, chromium, or iron ions are trivalent and should be powerful destabilizing agents, ie, a concentration of only 10^{-3} M for such ions should be sufficient for destabilization. Investigations on the stability of latices (17) have confirmed the prediction, but only for pH values below 3.5. At pH values above this limit the latex actually showed retained stability at aluminum ion concentrations of any value higher than 10^{-3} M. A diagram of the stability limits at various aluminum ion concentrations and pH values explains this result (Fig. 8) (17). At low pH values (<4) the ionic species is $[Al(H_2O)]^{3+}$ and the expected stability limiting concentration of 10^{-3} M is found. When the pH increases above that value a complex ion $[Al_6(OH)_{26}]^{4+}$ appears and the stability limit is strongly reduced. In fact the ion does not only act on the diffuse part of the electric double layer but adsorbs directly to the surface. In doing so the surface charge becomes reversed and the latex is stabilized by a positive charge rendering the aluminum ions without destabilizing effect. The latex is stable against enhanced concentration of aluminum ions. This effect has not been studied for emulsions per se but is expected to play a decisive role for these also.

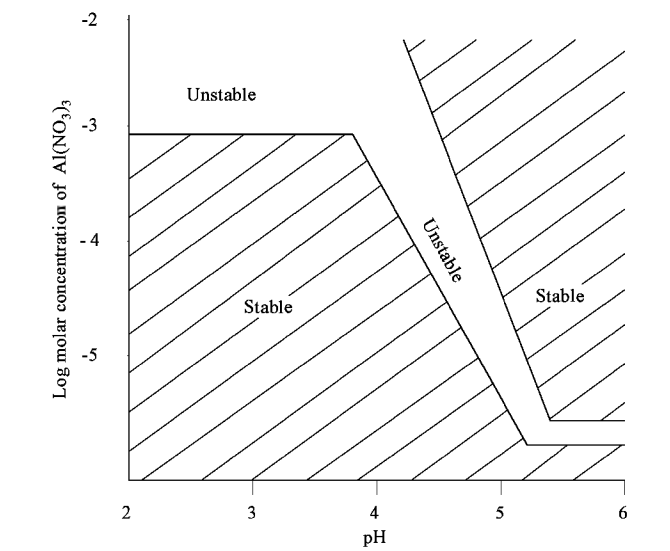


Fig. 8. The stability and instability regions of a polystyrene latex with added aluminum salt at different pH values (17).

POLYMER STABILIZATION

Polymers have so far been used comparatively less than the common surfactants to stabilize emulsions in spite of the fact that excellent stabilization by them can be achieved (18–20). Application probably has been limited because the adsorption of polymers to emulsion droplets has displayed some intricate phenomena; small changes in polymer structure or in solvent properties may lead to drastic changes in adsorption.

A polymer is adsorbed in the form of loops, tails, and trains (Fig. 9). Sufficiently long tails or loops provide stabilization. The action is extremely efficient (20,21); a single loop or tail gives a barrier of approximately 20 kT.

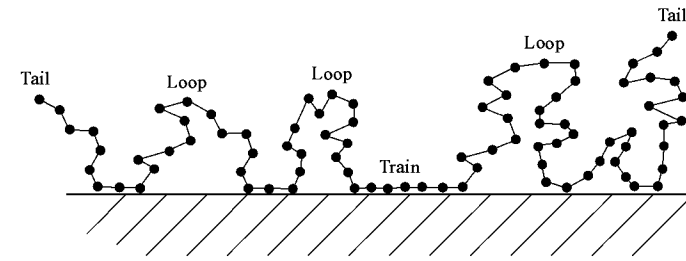


Fig. 9. A polymer is adsorbed to a surface in loops, tails, and trains.

However, polymer stabilization is sensitive to the properties of the environment. This has been described in an early calculation of polymer conformation at an interface (21). The fraction of adsorbed polymer was plotted as a function of its total adsorption energy (number of adsorbing groups times their individual adsorption energy) at constant molecular weight, showing the range for useful polymer stabilization to be extremely narrow. Adsorption energies lower than the optimal range (A, Fig. 10) give no adsorption of the polymer and no stabilization. At a higher adsorption energy (B, Fig. 10), the polymer adsorbs flat at the interface. Such an adsorption is also without stabilization effect because at short distances the van der Waals potential has already reached such large negative values that the potential well is too deep for the droplets to be deaggregated. Only in a limited range of adsorption energies, in which loops and tails are formed, does the polymer serve. In addition, the same phenomenon means that a minimum molecular weight is necessary to obtain stability.

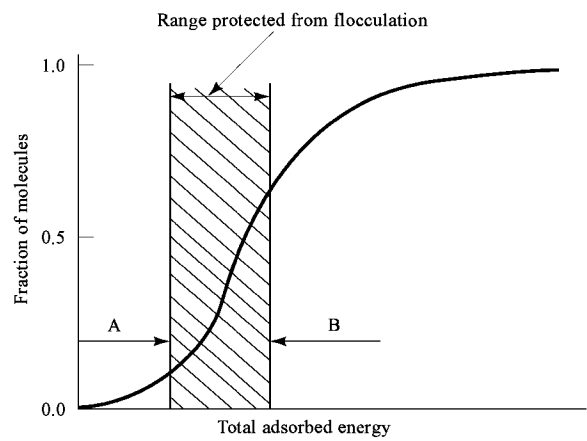


Fig. 10. Calculations of the adsorption of a polymer at an interface show pronounced sensitivity to the total adsorption energy, leaving only a limited range in which the polymer can serve as a stabilizer (21). See also text.

These problems may be overcome by using block copolymers. The polymer blocks are chosen to be selectively soluble in the aqueous and oil phase, respectively (see COPOLYMERS).

Stability of Three-Phase Emulsions

In the simplest emulsions just described, the final separation is into two liquid phases upon destabilization. The majority of emulsions are of this kind, but in some cases the emulsion is divided into more than two phases. One obvious reason for such a behavior is the presence of a material that does not dissolve in the oil or the water. One such case is the presence of solid particles, which is common in emulsions for food, pharmaceuticals, and cosmetics. Another less trivial reason is that the surfactant associates with the water and or the oil to form a colloidal structure that spontaneously separates from the two liquid phases. This colloidal structure may be an isotropic liquid or may be a semisolid phase, a liquid crystal, with long-range order.

In the case of emulsions with three liquids the presence of the third phase results in a reduction of the energy input for the emulsification process, whereas the emulsion with a liquid crystal as the third phase shows interesting stabilization mechanisms. Finally, the emulsion with added particles illustrates the importance of liquid–solid wetting for stability.

Liquid Third Phase. A third liquid with colloidal structure has been a known component in emulsions since the 1970s (22) for nonionic surfactants of the poly(ethylene glycol) alkylaryl ether type. It allows low energy emulsification (23) using the strong temperature dependence of the colloidal association structures in the water–surfactant–hydrocarbon systems.

At low temperature, nonionic surfactants are water-soluble but at high temperatures the surfactant's solubility in water is extremely small. At some intermediate temperature, the hydrophile–lipophile balance (HLB) temperature (24) or the phase inversion temperature (PIT) (22), a third isotropic liquid phase (25), appears between the oil and the water (Fig. 11). The emulsification is done at this temperature and the emulsifier is selected in the following manner. Equal amounts of the oil and the aqueous phases with all the components of the formulation pre-added are mixed with 4% of the emulsifiers to be tested in a series of samples. For the case of an o/w emulsion, the samples are left thermostated at 55°C to separate. The emulsifiers giving separation into three layers are then used for emulsification in order to find which one gives the most stable emulsion.

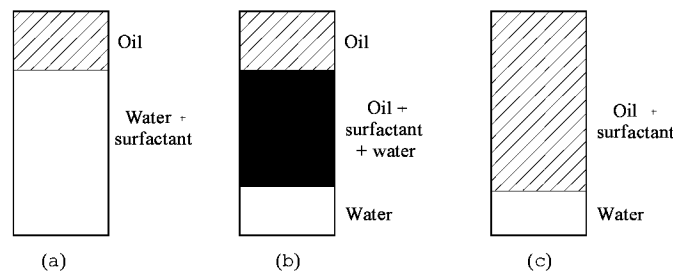


Fig. 11. In a system of water and hydrocarbon a nonionic emulsifier with a poly(ethylene glycol) chain as the polar part dissolves in the aqueous phase at low temperatures (a) and in the oil phase at high temperatures (c). At an intermediate temperature (b) three isotropic liquid phases may be found.

The emulsions are extremely unstable at the HLB temperature (22), and cooling to room temperature must be rapid. This can be achieved by spreading the emulsion in a thin layer on a cold plate or by letting it pass between cold rollers. For preparation of a stable o/w emulsion with sufficiently low o/w volume ratio, the following method is recommended. Before emulsification an amount corresponding to half the total emulsion is removed from the aqueous phase and cooled close to 0°C. The remaining part of the emulsion is emulsified at 55°C and stirred directly into the cooled aqueous part. The average droplet size of these emulsions is less than 1 μm with very little stirring. With these small droplets the stability is good and can be improved further by addition of small amounts of an ionic surfactant (≈0.05–0.1%).

Liquid Crystal Third Phase. In addition to micelles and microemulsion droplets, surfactants may form liquid crystals. A liquid crystal is a separate phase, which comes out of solution, not like the micelles or microemulsion droplets, which are microscopic entities within the solution.

The introduction of liquid crystals as stabilizing elements for emulsions occurred in 1969 when it was found that the sudden stabilization at emulsifier concentrations in excess of 2.5% of a water-*p*-xylene emulsion by a commercial octa(ethylene glycol) nonylphenyl ether was due to the formation of a liquid crystalline phase in the emulsion (26). Later investigations confirmed the strong stabilizing action of these structures (27).

The detection of liquid crystal is based primarily on anisotropic optical properties. This means that a sample of this phase looks radiant when viewed against a light source placed between crossed polarizers. An isotropic solution is black under such conditions (Fig. 12). Optical microscopy may also detect the liquid crystal in an emulsion. The liquid crystal is conspicuous from its radiance in polarized light (Fig. 13). The structure of the liquid crystalline phase is also most easily identified by optical microscopy. Lamellar liquid crystals have a pattern of oil streaks and Maltese crosses (Fig. 14a), whereas ones with hexagonal arrays of cylinders give a different optical pattern (Fig. 14b).

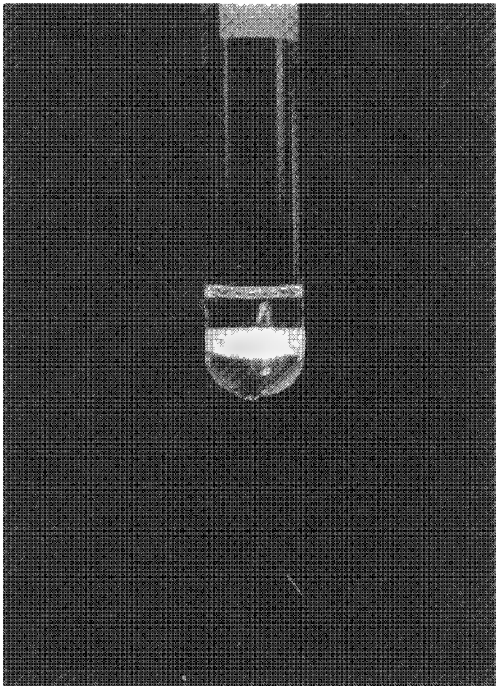


Fig. 12. A liquid crystal layer is radiant (middle) when placed between crossed polarizers and viewed against a light source.

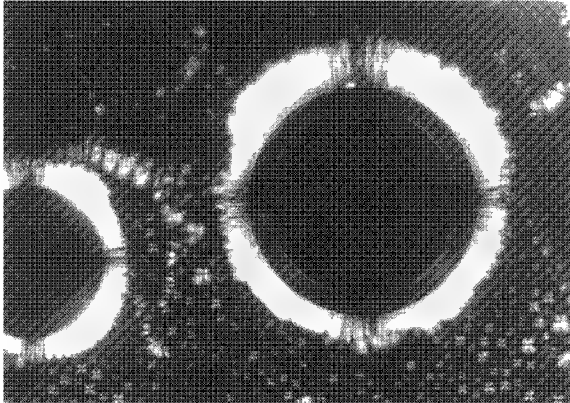


Fig. 13. Optical microscope shows the presence of a liquid crystal directly in the emulsion.

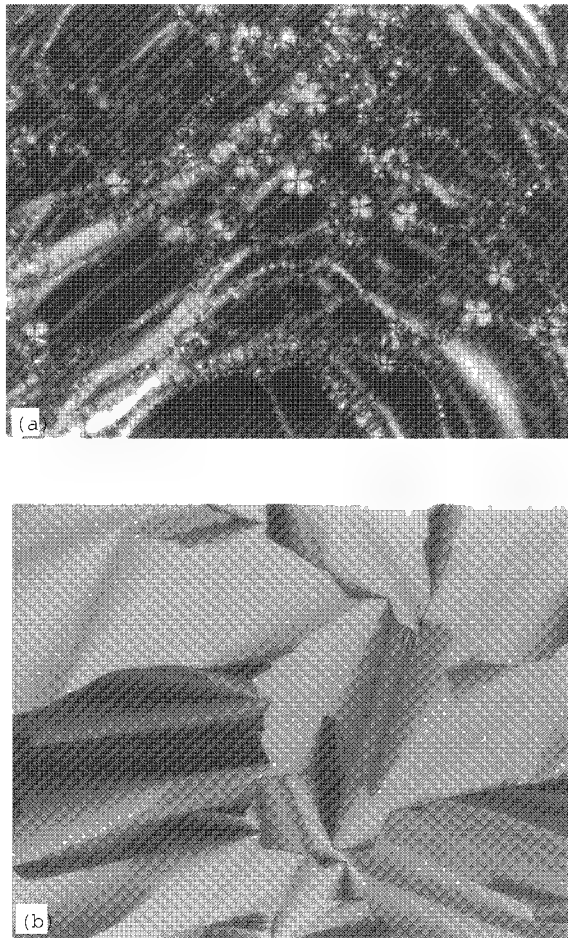


Fig. 14. A sample of a lamellar liquid crystal between crosses polarized in an optical microscope gives a pattern of "oily streaks" and Maltese crosses (a) while the liquid crystal consisting of an array of cylinders shows the characteristic sectional pattern (b).

Liquid crystals stabilize in several ways. The lamellar structure leads to a strong reduction of the van der Waals forces during the coalescence step. The mathematical treatment of this problem is fairly complex (28). A diagram of the van der Waals potential (Fig. 15) illustrates the phenomenon (29). Without the liquid crystalline phase, coalescence takes place over a thin liquid film in a distance range, where the slope of the van der Waals potential is steep, ie, there is a large van der Waals force. With the liquid crystal present, coalescence takes place over a thick film and the slope of the van der Waals potential is small. In addition, the liquid crystal is highly viscous, and two droplets separated by a viscous film of liquid crystal with only a small compressive force exhibit stability against coalescence. Finally, the network of liquid crystalline leaflets (30) hinders the free mobility of the emulsion droplets.

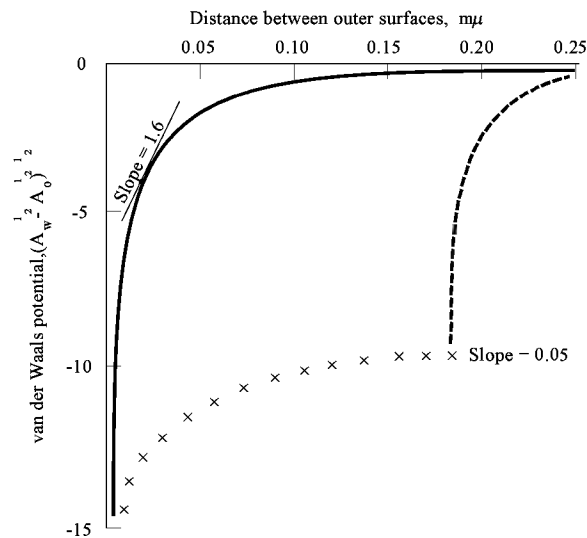


Fig. 15. The slope of the van der Waals potential vs interdroplet distance becomes extremely small in the coalescence step between two droplets covered by a lamellar liquid crystal.

The final factor influencing the stability of these three-phase emulsions is probably the most important one. Small changes in emulsifier concentration lead to drastic changes in the amounts of the three phases. As an example, consider the points A to C in Figure 16. At point A, with 2% emulsifier, 49% water, and 49% aqueous phase, 50% oil and 50% aqueous phase are the only phases present. At point B the emulsifier concentration has been increased to 4%. Now the oil phase constitutes 47% of the total and the aqueous phase is reduced to 29%; the remaining 24% is a liquid crystalline phase. The importance of these numbers is best perceived by a calculation of thickness of the protective layer of the emulsifier (point A) and of the liquid crystal (point B). The added surfactant, which at 2% would add a protective film of only 0.07 μm to emulsion droplets of 5 μm if all of it were adsorbed, has now been transformed to 24% of a viscous phase. This phase would form a very viscous film 0.85 μm thick. The protective coating is more than 10 times thicker than one from the surfactant alone because the thick viscous film contains only 7% emulsifier; the rest is 75% water and 18% oil. At point C, the aqueous phase has now disappeared, and the entire emulsion consists of 42.3% oil and 57.5% liquid crystalline phase. The stabilizing phase is now the principal part of the emulsion.

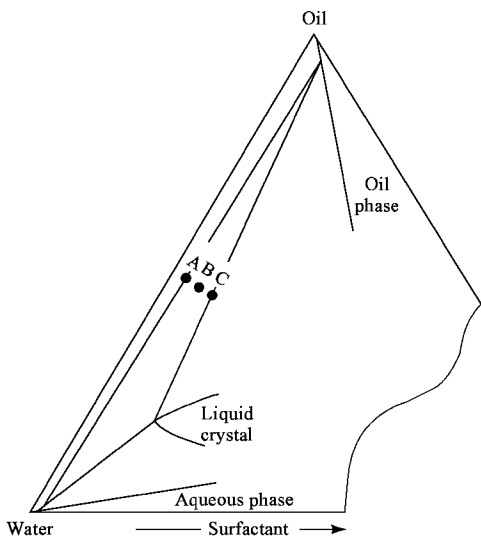


Fig. 16. An increase of the emulsifier concentration in an emulsion with a liquid crystal leads to a drastic increase of the amount of liquid crystal. See text.

One application of these systems is high internal ratio emulsions. In such an emulsion the dispersed phase occupies a large share of the total volume, resulting in an array of close-packed droplets. Flocculation is, hence, not avoidable; the structure often consists of a single aggregate of flocculated droplets. Protection against coalescence now becomes the primary consideration in emulsion formulation, liquid crystals being useful in preventing coalescence but not flocculation. Another application is concerned with spontaneous emulsification, a phenomenon which has been treated in a series of articles (31).

The conditions for surfactants to be useful to form liquid crystals exist when the cross-sectional areas of the polar group and the hydrocarbon chain are similar. This means that double-chain surfactants are eminently suited, and lecithin (qv) is a natural choice. Combinations of a monochain ionic surfactant with a long-chain carboxylic acid or alcohol yield lamellar liquid crystals at low concentrations, but suffer the disadvantage of the alcohol being too soluble in the oil phase. A combination of long-chain carboxylic acid plus an amine of equal chain length suffers less from this problem because of extensive ionization of both amphiphiles.

Two Liquids Plus a Solid. Solid particles may be used to stabilize an emulsion, avoiding the problem of simultaneous stabilization of both the oil drops of the emulsion and the solid particles of the suspension. The key factor for the use of particles as stabilizers is their location. If they are located at the interface between the two liquids, they will stabilize the emulsion, serving as a mechanical barrier to prevent the coalescence of the droplets (Fig. 17).

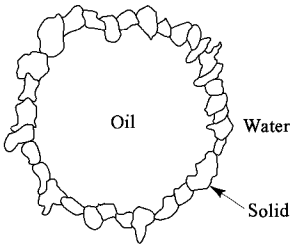


Fig. 17. Solid particles adsorbed at the oil-water interface serve as stabilizers for an emulsion.

The protection against coalescence is directly dependent on the energy to expel the particles from the interface. It depends on the contact angle and is easily calculated for the case of spherical particles located at the oil-water interface. The expression is equation 6 in which ΔE is the energy to expel a spherical particle with radius r from the interface into the phase by which it is predominantly wetted and toward which its contact angle is θ and γ_{o-w} is the interfacial tension between the oil and water phases. Experimental investigations (32) have shown 90° to be a practical maximum. Higher values probably lead to the expulsion of the particle into the continuous phase.

$$\Delta E = \pi r^2 \gamma_{o-w} (1 - \cos \theta)^2 \tag{6}$$

The wetting energies give repulsive forces exceeding that of the van der Waals attractive force under certain conditions. The combined van der Waals and wetting force is given by equation 7 in which h is the distance perpendicular to the interface that the particle has moved from its equilibrium position.

$$dE/dh = 2\pi \gamma_{o-w} (h - r \cos \theta) \tag{7}$$

Figure 18 is drawn for particles of 5 nm radius acting on a flat interface of 10^{-2} nm^2 . The total force is repulsive even for contact angles of 60° ; for 90° angles the repulsion is extremely strong. Hence, the stabilization is excellent, bearing in mind that thousands of such particles are adsorbed onto one emulsion droplet.

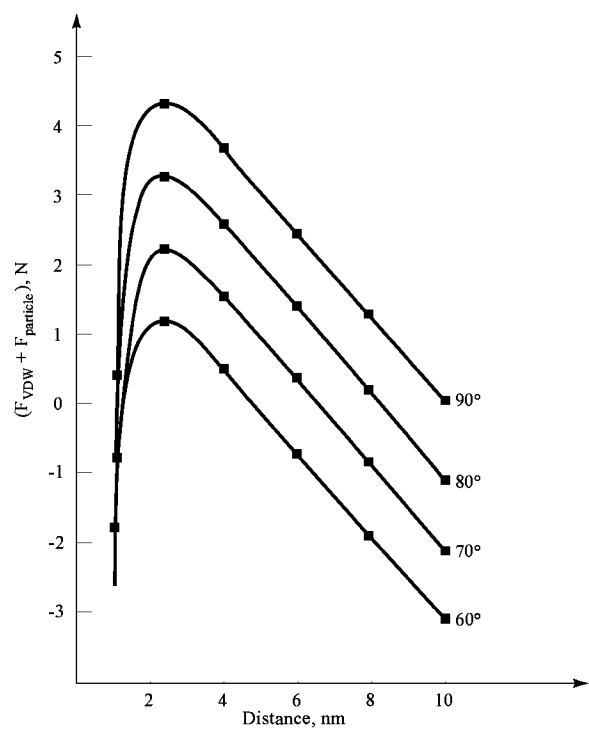


Fig. 18. The repulsion force from adsorbed particles is greater than the van der Waals force between flocculated emulsion droplets under certain conditions. Contact angles are noted on the curves.

The essential information from equation 6 and Figure 18 is that the protection energy from adsorbed solid particles is rapidly lost with reduced contact angle. When the contact angle is reduced, the particle is spontaneously moved toward the interior of the droplet and its complete expulsion into the droplet requires less energy. Figure 19 shows that the energy to displace one particle into the dispersed droplets from its location at the interface is reduced to half its value as the contact angle changes from 90° to 75°. An angle decrease to 60° leads to an energy value only one-fourth of that at 90°.

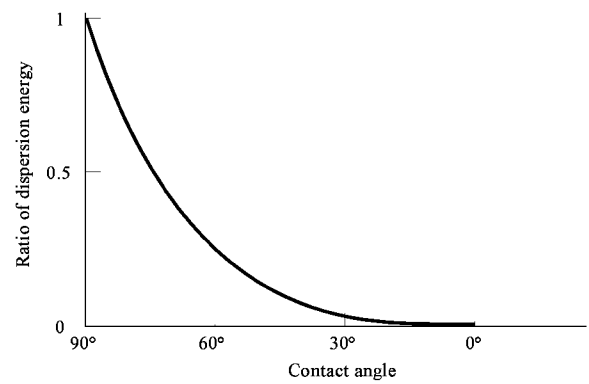


Fig. 19. The energy to displace a particle from the interface into one phase strongly depends on the contact angle.

These values make it evident that the contact angles must be observed rather carefully to obtain a value close to 90°. The following experimental procedure is practical to obtain good stability. A container with the particles is placed on the bottom of an oil-filled vessel with parallel walls, and a drop of the aqueous phase is placed on the powder. The shape of the droplet decides the action to be taken.

A droplet shape indicating a contact angle of 90° between the water and the solid material means that the solid particles are optimally useful to stabilize the emulsion with no further treatment. A contact angle smaller than 90° between the water and the solid particles means that the interfacial free energy is too great between the solid material and the oil. To increase the contact angle between the aqueous phase and the powder, an oil-soluble surface-active agent must be added. This surfactant should not be water-soluble at all and it should bind strongly to the solid surface. Such a surfactant binds strongly to the solid-liquid interface, and a low concentration in the oil phase is sufficient to reduce the interfacial free energy at the oil-solid interface. Use of as low a concentration as possible is essential because higher concentrations also lower the interfacial tension at the water-oil interface.

A reduction of the oil-water interfacial tension has a disadvantage because it makes the contact angle θ more sensitive to small differences between γ_{ws} and γ_{os} . After a certain concentration of surfactant in the oil phase has brought the contact angle to 90°, the process is repeated but with the surfactant added to the oil before the phases are brought into contact. If the water droplet does not spread and its contact angle is in excess of 90°, the surfactant is added to the aqueous phase.

For emulsions produced in high quantities or batched repeatedly, the surface of the solid particles should be changed through a chemical reaction. A small contact angle in the aqueous phase means that the surface of the solid particle must be made more hydrophobic. Esterification of polar groups on the particle surface by silanization or by similar processes is convenient. The other case, a large contact angle toward the aqueous phase, is corrected by making the particle surface more polar through oxidation.

BIBLIOGRAPHY

"Emulsions" in *ECT* 1st ed., Vol. 5, pp. 692-718, by W. C. Griffin, Atlas Powder Co.; in *ECT* 2nd ed., Vol. 8, pp. 117-154, by W. C. Griffin, Atlas Chemical Industries, Inc.; in *ECT* 3rd ed., Vol. 8, pp. 900-930, by W. C. Griffin, ICI United States Inc.

1. P. Becher, *Emulsions: Theory and Practice*, Reinhold, New York, 1965.
2. P. Sherman, *Emulsion Science*, Academic Press, Inc., New York, 1969.
3. S. E. Friberg, L. Mandell, and M. Larsson, *J. Coll. Interface Sci.* **29**, 1 (1969).

4.

International Union of Pure and Applied Chemistry, *Manual on Colloid and Surface Science*, Butterworths, London, 1972.

5.

M. M. Rieger in L. Lachman, H. A. Lieberman, and J. L. Kanig, eds., *Theory and Practice of Industrial Pharmacy*, 3rd ed., Lea & Febiger, Philadelphia, 1986, p. 502.

6.

A. F. Brodin, D. R. Kavaliunas, and S. G. Frank, *Acta Pharm. Suecica* **15**, 1 (1978).

7.

J. W. Frankenfeld, W. J. Asher, and N. N. Li, *Recent Developments in Separation Science*, Vol. 4, CRC Press, Boca Raton, Fla., 1978.

8.

K. Madani and S. E. Friberg, *Prog. Colloid Polym. Sci.* **65**, 164 (1978).

9.

P. Walstra in P. Becher, ed., *Encyclopedia of Emulsion Technology*, Marcel Dekker, 1983, p. 74.

10.

P. Walstra, *Chem. Eng. Sci.* **48**, 333 (1993).

11.

J. T. Davies, *Turbulence Phenomena*, Academic Press, Inc., New York, 1972, Chapt. 8–10.

12.

P. Walstra, *Neth. Milk Dairy J.* **29**, 279 (1975).

13.

P. Walstra in Ref. 9, p. 122.

14.

B. Derjaguin and L. D. Landau, *Acta Physicochim. USSR* **14**, 633 (1941).

15.

E. J. W. Verveij and J. T. G. Overbeek, *Theory of the Stability of Lyophobic Colloids*, Elsevier, Amsterdam, the Netherlands, 1949.

16.

W. Albers and J. T. G. Overbeek, *J. Colloid Sci.* **15**, 489 (1960).

17.

E. Matijević, *Pure Appl. Chem.* **53**, 2167 (1981).

18.

J. T. C. Boehm and J. Lyklema in A. L. Smith, ed., *Theory and Practice of Emulsions Technology*, Academic Press, Inc., New York, 1976, p. 23.

19.

D. E. Graham and M. C. Phillips in Ref. 18, p. 75.

20.

D. E. Napper, *Polymeric Stabilization of Colloidal Dispersion*, Academic Press, Inc., New York, 1983.

21.

E. J. Clayfield and E. C. Lumb, *Macromolecules* **1**, 133 (1968).

22.

K. Shinoda and H. Arai, *J. Phys. Chem.* **68**, 3485 (1964).

23.

T. J. Lin, *J. Soc. Cosmet. Chem.* **29**, 117 (1978).

24.

H. Saito and K. Shinoda, *J. Colloid Interface Sci.* **35**, 359 (1971).

25.

S. E. Friberg and I. Lapczynska, *Prog. Colloid Polym. Sci.* **56**, 16 (1975).

26.

S. E. Friberg, L. Mandell, and M. Larsson, *J. Colloid Interface Sci.* **29**, 155 (1969).

27.

S. E. Friberg and K. Larsson in *Advances in Liquid Crystals*, Vol. 2, Academic Press, Inc., New York, 1976, p. 173.

28.

P.-O. Jansson and S. E. Friberg, *Mol. Cryst. Liquid Cryst.* **34**, 75 (1976).

29.

S. E. Friberg, P.-O. Jansson, and E. Cederberg, *J. Colloid Interface Sci.* **55**, 614 (1976).

30.

B. W. Barry, *Adv. Colloid Interface Sci.* **5**, 37 (1975).

31.

M. J. Groves and H. S. Yalabik, *Pharm. Technol.* **2**, 21 (1977).

32.

J. H. Schulman and I. Leja, *Trans. Faraday Soc.* **50**, 598 (1954).

Stig E. Friberg
Steven Jones
Clarkson University

ENAMELS, PORCELAIN OR VITREOUS

Porcelain enamel is a glassy coating applied to various metal substrates through a high (usually >425°C) temperature fusion process (see also COATINGS) (1). The resultant chemically bonded glass–metal composite exhibits properties which reflect the chemical, physical, electrical, and aesthetic properties of the glass while combining the strength, ease of fabrication, and durability of the metal (see also COMPOSITE MATERIALS). Porcelain enamels are formulated to develop specific properties on metals or alloys.

The porcelain enamel is composed of various inorganic metal oxides fused between 1100 and 1400°C to form an alkali borosilicate glass (qv). The glass is rapidly quenched to produce either flakes by rapid chilling using water-cooled rollers, or glass granules by water quenching a narrow stream of molten glass. The resultant product is known as frit [65997-18-4]. An enameled article is produced by applying either wet or dry finely ground frit particles to the metal substrate. The enamel is fired to form one or more glassy layers to achieve a desired surface finish and property group for the article.

Porcelain enameling has been an important industrial process since the late nineteenth century in the United States. Modern appliances were developed utilizing the chemical and physical properties of the glass-on-metal coatings on cast iron (qv), sheet steel (qv), and, more recently, aluminum (see ALUMINUM AND ALUMINUM ALLOYS). The wide variety of colors and textures and the long-term durability of porcelain enamel surfaces have been attractive to appliance producers as well as users. Enamels have been used on a wide array of kitchen utensils, cooking devices such as ovens and range tops, sanitary ware such as bathtubs and lavatories, water heater tanks, specially built industrial chemical vessels, cast-iron piping, storage silos for agricultural and industrial products, as well as architectural interior and exterior surfaces. Porcelain enamel surfaces on architectural panels for domestic and industrial buildings have provided evidence of the long-term environmental durability of the surface. Even buildings dating from the 1940s have required only a minimum of maintenance. Controlled exposure experiments involving enamels at different geographic locations in the United States have shown only minimal color and surface degradation after 30 years of exposure near ocean and industrial locations (2).

Porcelain enamels are used in modern mass transit facilities because of inherent fire resistance as well as ease of maintenance and durability in the face of highly corrosive atmospheres resulting from vehicular traffic. Newer applications of porcelain enamels include electronic substrates (see ELECTRONIC MATERIALS), pyrolytic self-cleaning oven interiors, microwave oven interiors, outdoor cookers, and fireplace liners as well as wood-stove exteriors, writing boards for erasable markers, institutional surfaces such as bathroom stalls and hospital operating room walls that need to be easily disinfected, and subway car interiors and elevator walls that are durable and easy to clean. Porcelain enamel surfaces are frequently used in food contact applications. The glassy nature of porcelain enamels has a significant effect on encapsulating and minimizing any solubility of constituents. Enamels can also be formulated to have specific acid, alkali, and abrasion-resistant properties.

The early history of enameling is associated with the craft of enameling art. Porcelain enamel provides a low cost, extremely pleasing, nonporous protective and corrosion-resistant surface, using mass production techniques. There are many excellent summaries of the history of porcelain enameling and artistic uses of enamels. Porcelain enameling art is a continuing form of expression for many artists. Although enameling has been known for at least four to five thousand years, only in the twentieth century has the widespread use of enamels on base metals such as cast iron, sheet steels, and aluminum alloys flourished.

The Enameling Process

Melting. The frit-making process involves all the technology ranging from proper selection of raw materials through mixing and smelting to uniform production of frit. Box smelters, predating continuous smelters, are used for small quantity production. Electric smelting of porcelain enamel glasses has been carried out in the United States and abroad. The most prevalent method of making porcelain enamel glass is in continuous smelters, as shown in Figure 1. Continuous smelters may produce from 700 to 1200 kilograms of frit per hour. First, the glass is smelted from raw materials such as those listed in Table 1. Typical ranges of formulations are listed in Table 2. In a continuous furnace the thoroughly mixed raw batch is fed in at one end and molten glass flows out at the other end. Decomposition, gas evolution, reaction, and dissolution occur during smelting. Modern smelters, which are fired by natural gas and often pure oxygen, are equipped with pollution control equipment on the smelter exhaust system (4) (see AIR POLLUTION CONTROL METHODS; EXHAUST CONTROL, INDUSTRIAL).

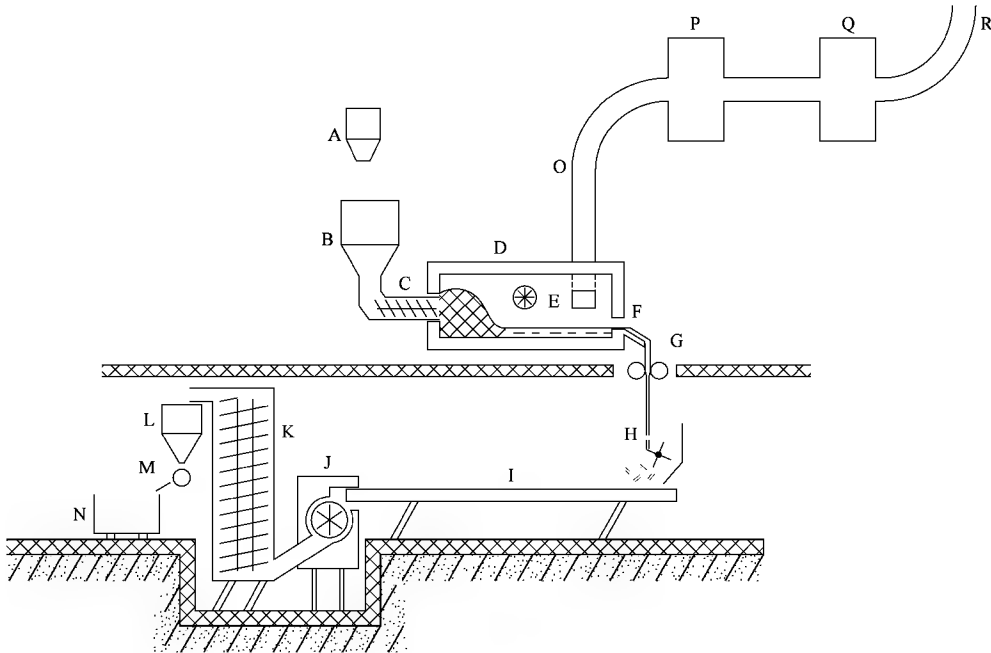


Fig. 1. Schematic of continuous smelter for porcelain enamel glass production where A is a raw batch hopper; B, batch hopper mixer; C, screw feeder; D, refractory furnace; E, burner for gas and air (oxygen); F, spout; G, water-cooled rollers for quenching glass; H, finger breakers to break glass ribbon; I, horizontal vibratory conveyor; J, hammer mill; K, vertical vibratory conveyor; L, surge bin; M, rotary magnetic separator; N, product container; O, exhaust gas flue; P, dry scrubber; Q, wet scrubber; and R, exhaust stack.

Table 1. Raw Materials for Frit Manufacture

Raw material	CAS Registry Number	Oxide content
alumina	[1344-28-1]	Al ₂ O ₃
antimony pentoxide	[1314-60-9]	Sb ₂ O ₅
antimony trioxide	[1309-64-4]	Sb ₃ O ₃
barium carbonate	[513-77-9]	BaO
borax decahydrate	[1303-96-4]	B ₂ O ₃ , Na ₂ O
borax pentahydrate	[11130-12-4]	B ₂ O ₃ , Na ₂ O
borax (anhydrous)	[1330-43-4]	B ₂ O ₃ , Na ₂ O
boric acid	[10043-35-3]	B ₂ O ₃
cobalt oxide	[1308-06-1]	Co ₂ O ₃
copper oxide	[1317-38-0]	CuO
feldspar ^a	[68476-25-5]	SiO ₂ , Al ₂ O ₃ , K ₂ O, Na ₂ O
litharge	[1317-36-8]	PbO
lithium carbonate	[554-13-2]	Li ₂ O
magnesite ^b	[13717-00-5]	MgO
manganese dioxide ^c	[1313-13-9]	MnO ₂
nepheline syenite ^d	[37244-96-5]	SiO ₂ , Al ₂ O ₃ , Na ₂ O
nickel oxide	[1313-99-1]	NiO
potassium carbonate	[584-08-7]	K ₂ O
potassium silicofluoride	[16871-90-2]	SiO ₂ , K ₂ O, F
quartz	[14808-60-7]	SiO ₂
rutile ^{a,b}	[1317-80-2]	TiO ₂
soda ash	[497-19-8]	Na ₂ O
sodium nitrate	[7631-99-4]	Na ₂ O, NO ₂
sodium silicofluoride	[16893-85-9]	Na ₂ O, SiO ₂ , F
sodium tripolyphosphate	[7758-29-4]	P ₂ O ₅ , Na ₂ O
spodumene ^e	[1302-37-0]	SiO ₂ , Al ₂ O ₃ , Li ₂ O
strontium carbonate ^c	[1633-05-2]	SrO
titanium dioxide	[13463-67-7]	TiO ₂
whiting ^f	[13397-26-7]	CaO
zinc oxide	[1314-13-2]	ZnO
monoammonium phosphate	[7722-76-1]	P ₂ O ₅
red iron oxide	[1309-37-1]	Fe ₂ O ₃
sodium fluoride	[7681-49-4]	Na ₂ O, F
vanadium pentoxide	[1314-62-1]	V ₂ O ₅

^a Also contains trace quantities of Fe₂O₃ and CaO.

^b Also contains trace quantities of SiO₂, Fe₂O₃, and Al₂O₃.

^c Also contains trace quantities of SiO₂ and R₂O₃ where R is an amphoteric oxide.

^d Also contains trace quantities of K₂O.

^e Also contains trace quantities of Fe₂O₃.

^f Also contains traces of MgO.

Table 2. Porcelain Enamel Frit Compositions, wt %^a

Oxide	Ground coats		Cover coats			Aluminum enamels	
	Sheet steel ^b	Cast iron ^c	Titania	Cast iron (leadless)	Cast iron (Pb–Sb)	Leadless ^d	Lead-bearing
SiO ₂	28–68	50–60	38–57	40–50	40–50	30–40	30–40
B ₂ O ₃	6–27	10–15	13–20	14–20	2–6	0–3	0–3
Na ₂ O	6–24	5–10	6–13	6–10	15–20	15–20	10–18
K ₂ O	0–8	3–8	2–11	6–10	0–2	7–12	5–10
Li ₂ O	0–7		0–2	0–2	0–2	2–5	2–5
CaO	0–15	0–3			3–8		
BaO	0–16				0–5	2–6	2–6
Al ₂ O ₃	0–12		0–5	0–3	2–6		
P ₂ O ₅	0–4		0–4	0–2		2–5	2–5
ZrO ₂	0–13	1–3	0–6	0–2	0–1		
MgO	0–1		0–2	0–2			
TiO ₂	0–10		14–25	15–20	5–10	12–20	12–20
ZnO	0–4		0–3	0–1			
Sb ₂ O ₅					5–10	2–5	2–5
PbO		2–6			5–12		10–20
F ₂	0–10		0–10	3–8	2–6		

NO ₂	0–7	0–4	3–5	3–5	3–6	3–6
^a	Lead, Pb; nickel, Ni; and barium, Ba, may be restricted or absent in some formulations because of environmental regulations (3).					
^b	Also contains 0–4 wt % Co ₂ O ₃ ; 0–5, NiO; 0–4, Fe ₂ O ₃ ; and 0–2, CuO.					
^c	Light base coat.					
^d	Also contains 3–8 wt % V ₂ O ₅ .					

After the molten glass has become a homogeneous liquid, it is poured in a thin stream into water or between closely spaced, water-cooled rotating metal rollers. This quenched glass, termed frit, is a friable material easily reduced to small-sized particles by a ball-milling operation. Ball-milling the glass frit into small-sized particles can be carried out whether the frit is to be used wet or dry (see [SIZE REDUCTION](#)).

Grinding. Ball-mill grinding of the glass is accomplished using porcelain or high alumina balls 1.5–5.0 cm in diameter. Mill size, speed, and charge, and ball size and charge are important parameters for the determination of the milling time required for optimum size distribution in prepared slip. Ground-coat enamels are ball-milled to a fineness of 95% of solids smaller than 74 μm (<200 mesh). Cover coats are ground more finely, to 98% of solids finer than 44 μm (325 mesh). Dry powders are used for dry process cast-iron enameling and for electrostatic application on sheet steel (see [COATING PROCESSES, POWDER TECHNOLOGY](#)). Dry powders are also prepared and marketed for the subsequent preparation of slurries and slips used in wet-process application techniques.

For conventional wet processing of sheet steel, the porcelain enamel frit is ball-milled using clay, certain electrolytes, and water to form a stable suspension. This clay-supported slurry of small particles is called the slip and has the consistency of a heavy cream. The ingredients of the mill batch are carefully controlled. The amount and purity of all materials in the mill, including the clay and water, affect the rheological character of the slip as well as a number of the properties of the fired enamel such as chemical resistance, reflectance, gloss, color, and abrasion resistance.

Application. Ground coats are applied to metal by spraying, dipping, or draining as well as electrostatically. Large articles (washing machine tubs and stove chassis) may be coated only by spraying or flow coating the slip, whereas small articles may be dipped into a tank of ground-coat slip. Cover coats may be applied by similar methods, by electrostatic spraying (wet and dry powder), or by electrophoresis (qv) (see [COATING PROCESSES](#)).

Since 1975, one-coat or two-coat dry electrostatic application of dry frit particles has been a significant methodology in the appliance industry worldwide. Nearly all ranges in the United States are produced using this process technique. Dry ball-milled frit particles are coated with a silicone surface treatment, suspended in a fluidized bed, and delivered to spray guns that impart a negative electric charge to each particle. The metal article to be coated is at ground potential.

The glass particles without the coating may have a resistivity of 10⁶ ohm·cm; using the silicone coating, the glass particles can achieve a resistivity as great as 10¹⁵ ohm·cm (5). This greater resistivity is necessary to enable the charged particle to adhere to the metal substrate. The resistivity under ambient humidity conditions controls the rate of charge decay and therefore the adherence of the glass particles to the substrate. An overall resistivity of 10¹⁴ ohm·cm permits a coating thickness of 100 to 200 μm. The coating thickness is self-limiting and thickness uniformity is excellent even on curved surfaces. This dry process obviates the need for drying and intermediate firing for two-coat articles resulting in substantial savings in labor, materials usage, and energy costs (6), as compared to the conventional wet-process porcelain enameling of sheet steel. The dry system also eliminates the need for clay suspending agents that increase the refractoriness of the wet enamel systems.

The dry powder process has several additional advantages over the wet process. For example, much less waste of enamel occurs because the dry over-spray is airborne and recycled in a closed system. No-pickle ground coats have broadened the application of both wet-process and dry-process systems. These enamels are applied over cleaned-only metal. Thus the problems of disposing of pickling acid wastes containing iron sulfates and nickel wastes are eliminated (see [METAL SURFACE TREATMENTS](#)) (7).

Prior to the development of the dry electrostatic process for sheet-steel enameling, the one-coat, direct-on wet process for cover coats had been adapted by a large sector of the industrial sheet-steel enamellers. This process requires the use of extra low carbon (decarburized) steel so that carbon oxide gas defects are eliminated. Also, a special pickling process is required to assure good adherence of the enamel. This process involves an accelerated etching of the metal by the addition of ferric sulfate in the sulfuric acid pickling solution, as well as subsequent heavy nickel flash deposits. Cover enamels are applied by flow-coating, spraying, dip-draining, or electrophoresis. A process flow diagram for sheet-steel enameling is shown in Figure 2.

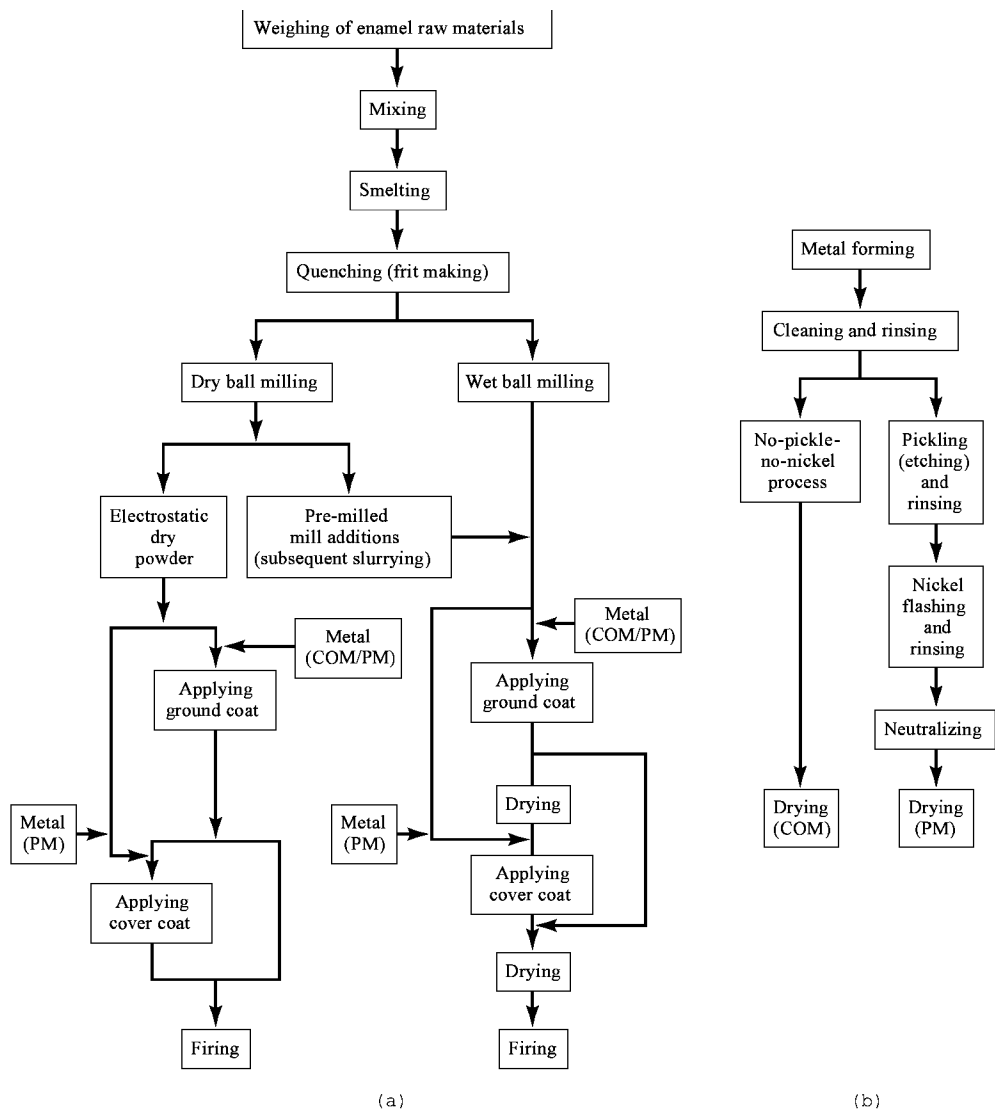


Fig. 2. Process flow diagram for sheet-steel enameling where COM indicates clean-only metal and PM, pickled metal, for (a) enamel preparation and application and (b) metal preparation.

Electrostatic wet enamel application, developed in the 1960s, is a minor commercial process in the United States. However, the technology is undergoing continuing development and improvement in process equipment. Disks, bells, and spray guns having high (100 kV) intensity d-c fields may be used to atomize and charge the enamel slip as it leaves the application apparatus. The negatively charged enamel particles are attracted to the surface of the metal article, which is at ground potential. This method of coating, which is usually automated, produces a uniform coating thickness and good coverage around corners (wrap-around) (8).

Electrophoretic application, also employed commercially on a limited basis, provides a dense uniform coating and excellent edge coverage. In this process the sheet-metal article is the positive electrode. Both the positive and negative electrodes are located in the enamel bath. The metal article, when subjected to a potential of 50 V and 50–300 A m² in the enamel slip, attracts the charged frit particles which then pack into a tightly adhering coating layer (9). This process has been automated in Europe for appliances. In the United States it has been adapted for electronic substrates (see ELECTRONIC COATINGS).

Dry-process cast-iron enameling involves the application of the cover coat by dusting or dredging the dry glass powder (using a long-handled vibrating sifter) onto the heated cast-iron article. The cast iron, which has been previously ground coated, is removed from the furnace (870°C and higher) for the dusting operation and then is returned to the furnace to complete the fusion and maturation of the glass. Generally, two or more dusting and heating cycles are required for the development of an adequately thick and uniform coating (10).

Process flow diagrams for cast-iron and aluminum enameling are shown in Figures 3 and 4.

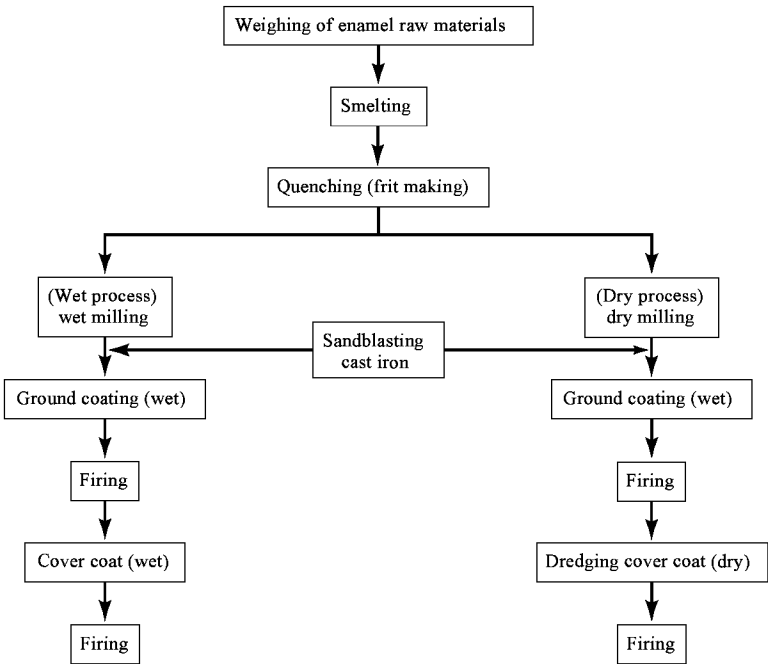


Fig. 3. Flow diagram of cast-iron enameling.

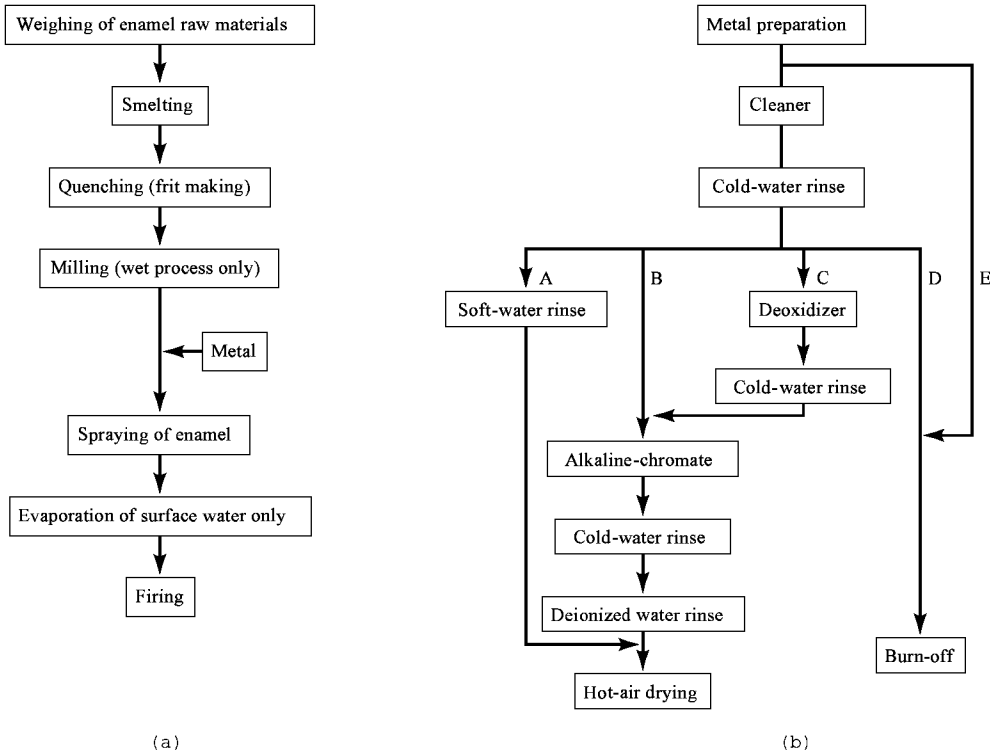


Fig. 4. Process flow diagram for aluminum enameling showing (a) enamel preparation and application, and (b) metal preparation, where the cleaning processes A, B, and C represent primarily sheet; D, primarily castings; and E, aluminized steel (11).

Metals for Enameling.

Steel. Sheet steel is usually bought in precut sheets or coils for subsequent stamping and pressing into shapes. The steels are chemically formulated to make them suitable for the fabrication and porcelain enameling operations. The ASTM (12) has classified enameling-grade steels into Type I and Type II (A or B) as well as specifying various qualities such as commercial, drawing, and drawing quality special killed. Type I has an extremely low carbon level, commonly produced by decarburizing in an open-coil process. This material is suitable for direct cover coat enameling practice. The less expensive Type II has moderately low carbon and is suitable for ground coat enameling operations. Steels having a II-A designation are intended for use where sag resistance during firing is of prime importance. The designation of II-B is for steels requiring good formability. Table 3 (13) lists the general chemical requirements for the three classes of steels.

Table 3. Sheet-Steel Chemical Composition,^{a,b} wt %

Element	Type I	Type II	
		A	B
carbon	0.008 ^c	0.04	0.08
manganese	0.60	0.15	0.50
phosphorus		0.015	0.015
sulfur ^d	0.040	0.040	0.040

^a Ref. 13.
^b Values given are maximum.

^c Cast analysis of carbon is not appropriate for Type I. Sheet products analysis is appropriate for checking proper type of material. Extremely low carbon levels can be checked accurately using carbon-combustion chromatographic-type analyzers.

^d Drawing quality has 0.035 wt % max.

The basic steel types are undergoing gradual modifications to adapt the steels to the continuous casting process. This has led to changes in the minor constituents of steel such as boron, nitrogen, titanium, and other alloying elements. Titanium is used to tie up the carbon as titanium carbide [12070-08-5], TiC, and the boron nitride [10043-11-5], BN, and titanium nitride [25583-20-4], TiN, provide void sites for hydrogen accumulation after firing (see CARBIDES, NITRIDES) (14).

Cast Iron. Cast irons for enameling contain between 2.8 and 3.7 wt % carbon; the more usual content is between 3.25 and 3.6 wt % (15). The carbon is usually found in two forms: graphitic carbon and combined carbon. Some additional elements in the cast iron are silicon, manganese, sulfur, and phosphorus. The cast iron known as gray cast iron is the most widely used for enameling purposes. Before enameling, a casting must be cleaned, usually by abrasively blasting the surface with sand or steel shot.

Aluminum. The most commonly used aluminum alloys for enameling are 3003 P.E. grade and 6061 P.E. grade. Other alloys are 1100 (commercially pure aluminum) 7104 (C74S) heat treatable, 43 casting alloy, and 344 and 356 casting alloys. It is important that the magnesium and copper contents of the alloys be kept to a minimum; otherwise, spalling of the enamel can occur (16).

Other Metals. Metals such as the austenitic series, Types 301–347, and the ferritic series, Types 409–446, of stainless steels may be enameled, as well as a number of other alloys (17). The metal preparation usually consists of degreasing and grit blasting. Copper, gold, and silver are also enameled. These metals are usually prepared for application by degreasing. Copper is pickled using either a nitric acid [7697-37-2] or a sulfuric acid [7664-93-9] solution, followed by a dilute nitric acid dip. Silver may be pickled in hot dilute sulfuric acid followed by a dip in a nitric acid solution (18).

Metal Preparation. Sheet-steel parts are formed by stamping, bending, and shearing. Many parts require welding (qv), which needs to be carried out in a uniform, smooth manner so that the welded joint can be enameled without defects. Cast-iron parts are formed by the usual cast-iron foundry methods; however, additional care is given to prevent contamination of the surface. Surface contamination causes defects in the enamel, particularly blisters and bubbles. Aluminum metal can be formed in sheets, extrusions, and as castings.

Enameling cannot be successful unless the metal is thoroughly cleaned and kept clean until the final coat is fired. Simply touching the surface with a hand can cause defects. Cast iron, thick steel parts, and aluminum castings may be sandblasted. The thickness precludes the danger of deformation resulting from metal loss. Sand, silicon carbide, and steel grit are satisfactory abrasives (qv). Products made from thin sheet material are most satisfactorily and most economically cleaned by chemical methods that require alkali and soap solutions to remove grease and dirt, and acid solutions to remove oxidized metal.

Steel. Commercial enameling metal pretreatment for sheet steel has undergone significant changes since the 1970s. The introduction of the no-pickle enamels for a variety of ground-coat applications has significantly reduced the use of acid etching and nickel flashing. Clean-only systems for grease and grime removal are widely used by appliance manufacturers, reducing costs of in-plant processing and the problems of waste disposal. Two-coat–one-fire enameling processes have also reduced the need for heavy-metal etching and nickel flashing (nickel replacement) for direct-on enamel application to decarburized steels.

Chemical treatment in industrial metal preparation can be carried out with continuous cleaning and pickling equipment in either an automated or a manual batch operation. The latter is used for smaller sized substrates. The fabricated sheet-steel articles are supported on continuously moving racks or hangers which first pass through cleaning stages. For the clean-only systems, this is all that is required. For the processes needing pickled steel, the substrate is then subjected to pickling acids, nickel flashing, and neutralizing treatments. Spray-pickling equipment is designed for rapid and easy treatment of large substrates and lends itself to continuous drying and transfer of ware.

Compositions of the cleaning solutions are dependent on the type of oil, grease, and grime to be removed, including the type and amount of drawing compounds used in the metal-forming operations. Vegetable oils (qv) may be saponified and removed by alkalies alone, but mineral oils must be removed using soap (qv). A well-balanced cleaning compound contains an alkali, an alkali salt (which acts as a buffer material to maintain an approximately constant pH), and a soap. Commercial cleaning products are available and are usually used at a strength of ca 45 g L (6 oz gal) of water at or near boiling temperatures.

After being cleaned, ware that is to be pickled is immersed successively in one or more tanks of water at 80–95°C and then transferred to the acid pickling solution. The pickling solution of 6–8% sulfuric acid is contained in a stainless steel tank or, alternatively, a lead-lined wooden tank at 60–65°C. The metal remains in the tank until all oxide scale is removed, usually 5–15 minutes. Rinse water, which removes acid salts, is maintained at 80–90°C and rinsing time is usually 3 minutes. Plating galvanically or flashing a thin film of nickel on the iron after rinsing retards oxidation and enhances the enamel bond. The flashing solution contains 7.5 g L (1 oz gal) of nickel salt, such as nickel sulfate [7786-81-4], NiSO₄·6H₂O, or the equivalent. The pH is maintained between 3.0 and 3.6 and the temperature at 75°C. The average time of immersion is 4–6 minutes and the deposit is usually 0.45–1.30 g m² (0.04–0.12 g ft²). A stainless steel tank is usually employed as the container.

For the cover-coat direct-on process, a ferric sulfate [10028-22-5], Fe₂(SO₄)₃, etch is included in the metal pretreatment for rapid metal removal. It is designed to remove ca 20 g m² (2 g ft²) of iron from the sheet metal surface. Hydrogen peroxide [7722-84-1], H₂O₂, is added intermittently to a 1% ferric sulfate solution to reoxidize ferrous sulfate [7720-78-7], FeSO₄, to ferric sulfate.

After being removed from the nickel bath, the ware is dipped into a hot or cold water rinse, quickly removed, and then transferred to a neutralizing bath where the last traces of acid are removed. Neutralizing using a solution of sodium carbonate (soda ash) and borax is common.

After being removed from the neutralizing solution, the ware is transferred to a dryer maintained at about 110–120°C that has good air circulation. This ensures quick and complete drying without rusting of the metal. After being dried, the sheet ware is ready for application of enamel.

Cast Iron. Cast iron is prepared for enameling by sand and steel shot blasting. The surface of the cast iron as it comes from the foundry has a thin hard skin which must be removed to achieve good enameled surface quality. The blasting operation also smoothes the surface, edges, and gates, and removes residual molding sand. Blasting is an operation that causes the abrasive grains to impact on the metal surface. Air-pressure blasting and centrifugally throwing the abrasive are common processes.

Aluminum. For aluminum alloys such as 1100 and 3003 only a cleaning operation is necessary before application of enamel. For alloy 6061, deoxidation and alkaline-chromate pretreatment processes may also be used to improve both the cleanliness of the surface and the development of enamel adherence. The spent chromium-containing solutions can present disposal problems and should be handled in accordance with existing regulations (19).

Firing of Enamels. Firing can be carried out in intermittent box-type furnaces or continuous furnaces (see FURNACES, ELECTRIC; FURNACES, FUEL FIRED). The dryer and the furnace form one continuous unit or function as separate units in the continuous firing process. Most industrial furnaces are fiber-lined (low thermal mass), which lowers cost and downtime between firing schedules. In a continuous tunnel-type furnace, each piece progresses through the furnace on a conveyer supported on alloy firing racks designed to withstand the repeated thermal cycles. Very efficient U-type furnaces, in which the outgoing ware preheats the incoming ware, have been built (20). Gas-fired furnaces have alloy muffles to separate the combustion gases from the furnace atmosphere. Electrically fired furnaces have special alloy resistance-heating elements on the surface of the refractory lining (see REFRACTORIES). Gradual heating and cooling of the ware is more easily achieved using continuous firing than by using intermittent firing. In the latter process, a batch of ware is loaded on a large rack of special support tools and is introduced into the box furnace.

Enamels are fired to mature coatings in a matter of minutes. The firing time in the hot zone of a continuous furnace is about 3–6 minutes at 750–850°C. Cover-coat firing is generally carried out at similar temperatures, especially when two-coat–one-fire application processes are used. For two-coat enameling, it is preferable to fire the ground coat about 30°C higher than the subsequent cover-coating, in order to minimize haidining in the cover coat and pull-through of the ground coat.

Enamel firing temperature is dictated by the coating composition, metal thickness, and the type of metal used. Enamels for aluminum are fired at

510–540°C, whereas those for high temperature alloys (qv) are fired up to 930°C.

The industrial porcelain enameling process can be automated to a great degree, from metal pretreatment (pickling) through the firing process. Flow-coating, electrostatic dry powder spraying, or electrophoretic deposition can be machine programmed. Robot sprayers controlled by a computerized sensing mechanism can coat complex shapes, and flow-coaters can be programmed to uniformly coat complicated shapes such as dishwasher cavities and washing machine baskets. Automated coating and firing of cast iron has also been achieved.

The level of moisture in the furnace atmosphere is also of importance in the development of good fired surface appearance. The range of moisture normally associated with good enameling practice is 1–2 vol % of moisture (21). Levels below 1 vol % may result in reduced gloss of the glass surface; levels above 2 vol % may result in blistering or a scummy surface. Winter conditions usually cause dry furnace atmospheres, and summer conditions may result in excessive moisture if the furnace is not properly vented.

Energy Requirements. The energy needed to heat 0.45 kg of enamel ware and tooling to the firing temperature ranges from 400–1000 kJ/kg (300–450 Btu/lb) and depends on such factors as furnace loading, tool weight, and, in gas-fired furnaces, flue-gas losses. Advances in furnace design, firing systems, and low thermal mass insulating materials have reduced the total energy needed over the years. Thus overall energy requirements for enameling have dropped as much as 50% since the 1970s. Technologies such as electrostatic frit powder application, various two-coat–one-fire processes (eg, dry-over-dry, dry-over-wet, and wet-over-wet), as well as reductions in the metal pretreatment requirements have also contributed (22).

Composition

Porcelain enamels are basically alkali borosilicate glasses. These enamels are complex, however, because of the large number and types of oxides which are needed to develop proper adherence and functional properties (23,24). Network-forming ingredients and modifiers are used as in normal glass (qv) making practice. The principal network formers are SiO₂, B₂O₃, and P₂O₅. Modifiers include the alkali metal oxides (Na₂O, K₂O, and Li₂O) and alkaline-earth metal oxides (CaO, BaO, and SrO). Other common oxides include Al₂O₃, MgO, ZrO₂, ZnO, TiO₂, Sb₂O₃, and the halide F. Transition-metal oxides such as Fe₂O₃, CoO, NiO, CuO, and MnO₂ are used for adherence to the sheet-steel substrate and color development in ground coats. Continuous-cleaning (catalytic) oven enamels have high percentages of the transition-metal oxides to achieve cleaning effectiveness. Less frequently used oxides of cerium, cadmium, lead, and tin are used in sheet-steel, cast-iron, and aluminum enamels.

Enamels used on cast iron and aluminum have traditionally been composed of SiO₂, B₂O₃, P₂O₅, and PbO. The lead oxide produces good surface quality, fusibility, and acid resistance when properly formulated with other oxides. More recently some nonlead-bearing compositions have been developed for both cast-iron and aluminum metals. Glasses containing lead oxide are not recommended for food contact surfaces.

Typical ranges of enamel compositions are listed in Table 2. Raw materials (Table 1) for the glass batch include minerals, such as feldspars and quartz, because these are inexpensive sources of SiO₂ and Al₂O₃ (see CLAYS). The batch composition for cover coats is comprised primarily of manufactured chemicals of known, controlled levels of purity to maintain reproducible, clean colors.

The silica content of a glass has a significant effect on the chemical and mechanical properties of the porcelain enamel. Increasing silica content is generally associated with increasing acid resistance and lowered thermal expansion. Increasing alkali content, particularly Na₂O, reduces acid resistance and increases the thermal expansion. Various modifiers such as Al₂O₃ (minor amounts) and ZrO₂ increase the alkali resistance, whereas TiO₂, generally an opacifier if in crystal form, improves the acid resistance. Self-opacified cover coats usually have titania as the crystalline phase; however, zirconia and antimony oxide are still used for some nonsheet-steel applications.

Porcelain enamels meet a variety of performance characteristics required for different applications. The common characteristics of all enamels include good adherence to the substrate and good thermal expansion fit to the metal. Specific properties depend on usage; for example, acid and alkali resistance, hot water resistance, abrasion resistance, thermal shock resistance, high gloss, high reflectance, specific color, heat resistance, and cleanability.

Titanium Dioxide. The recrystallization of titanium dioxide in a cover-coat glass is very important to the development of thin, highly opaque finish coats. Titania, TiO₂, is the primary opacifying agent for white finish coats. Two polymorphic forms of titania, anatase and rutile, may be present in the enamel (see TITANIUM COMPOUNDS, INORGANIC). Anatase is preferred because anatase crystals are present in the size range (0.1–0.2 μm) for maximum reflectance, and therefore generate the most desirable bluish white color. Smaller pigment particles provide too much light scattering and give a more undesirable bluish color; larger particles give a more cream-white color. Moreover, in contrast to rutile particles, anatase crystals do not grow or change size depending on the firing temperature.

When the porcelain enamel glass is smelted, the titanium dioxide is dissolved completely in the glass and exhibits a complete transparency. Upon subsequent reheating of the glass, nucleation sites within the glass matrix and at the surface of the glass particle initiate recrystallization of the titanium dioxide. Control of the thermal history of the glass reheating affects the subsequent crystal growth and size development. Color stability of titania enamels can be obtained by adjusting the composition of the glass so that anatase crystallizes predominantly over a wide temperature range. The effect of P₂O₅ content on the relative amounts of anatase and rutile crystallizing from a titania enamel has been shown to be important (25). Additionally, other oxides are known to have an effect on the anatase to rutile inversion. This inversion is thought to be an irreversible change from anatase to rutile crystal structure between 700 and 815°C (1300 and 1500°F). The anatase form of the titania crystal is preferred because of color and the general stability of the form which can be developed through the glass formulation. Figure 5 shows a typical cross-section of a titania-opacified direct-on enamel and metal interface.

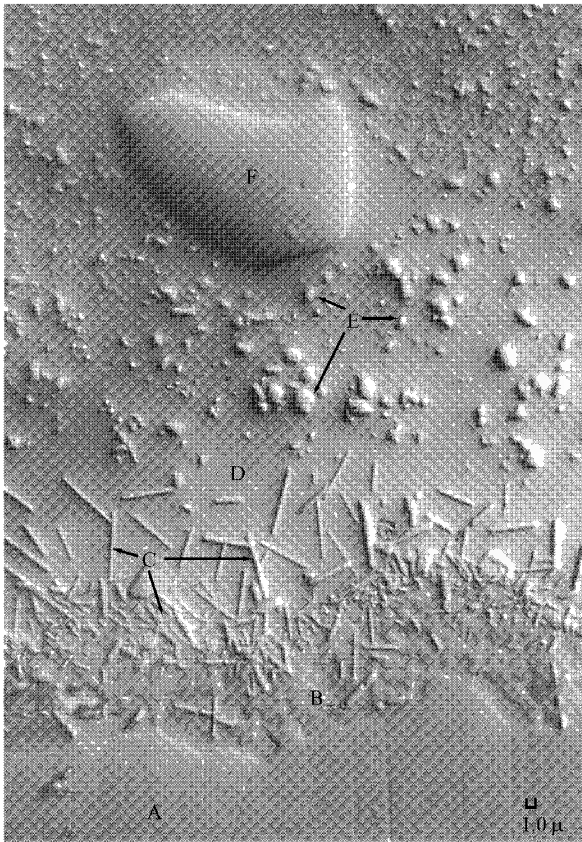


Fig. 5. A 90° polished cross section of a production white titania enamel, with the microstructure showing the interface between steel and direct-on enamel as observed by reflected light micrography at 3500 × magnification using Nomarski Interface Contrast (oil immersion). A is a steel substrate; B, complex interface phases including an iron–nickel alloy; C, iron titanate crystals; D, glassy matrix; E, anatase, TiO₂, crystals; and F, quartz particle.

Courtesy of Ferro Corp.

Properties

Thermal Fit and Residual Stresses. Thermal expansion measurements are typically carried out on dilatometric equipment consisting of a fused quartz pusher rod inside a tube of the same material (26). A bar specimen of the frit is prepared from crushed material of about 2000 μm (10 mesh) by reheating in a refractory or metal boat, annealing, and removing from the mold. A small furnace having a programmed temperature rise heats the tube containing the frit bar sample to a preset temperature or automatically shuts off when the dilatometric softening point is reached. The displacement of the pusher rod is automatically recorded, allowing a direct calculation of percentage of linear thermal expansion. Appropriate reference standards are used to calibrate the testing unit.

The expansion coefficient of the metal is usually constant over the entire range of the expansion of the glass, but the linear thermal expansion coefficient of the glass gradually increases with increasing temperature, accompanied by a noticeable change at the transformation temperature (*T_g*). The expansion of the bar appears to reverse after the softening point is reached because of sag. On the thermal expansion curves this is known as the dilatometric softening point. The glass bar becomes sufficiently fluid so that it cannot resist the spring pressure of the push rod and rapidly deforms.

A porcelain enamel glass becomes less viscous as the temperature is increased during firing. Above the softening point, the enamel is relatively fluid and conforms to the metal surface. As the porcelain enamel is cooled from the firing temperature of 750–800°C to the softening point, the fluid glass does not retain stress. However, as cooling proceeds below the softening point, the expansion (or contraction) of the coating exceeds that of the steel, and tensile stresses begin to develop in the coating, as shown in Figure 6. On further cooling, the stress increases until the temperature at which the expansion of the glass equals that of the metal. With still further cooling, the coefficient of expansion of the glass is less than that of the metal, coating tensile stresses decrease, and compressive stresses develop and are retained at room temperature.

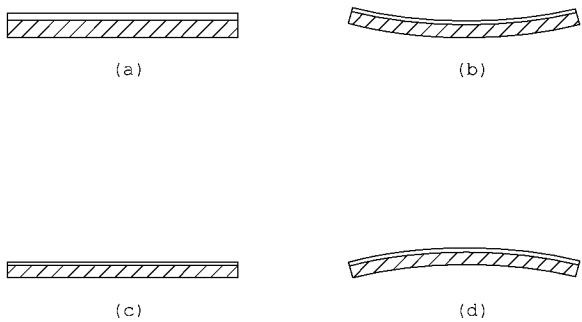


Fig. 6. Cooling cycle, where □ is the enamel layer and ▨ is steel, at (a) the upper no-strain point, where the enamel is molten, with no stress; (b) after solidification of the enamel, where the porcelain is in tension; (c) the lower no-strain point, where no stress is evident; and (d) at room temperature, where the porcelain enamel is in compression (27).

Courtesy of Ferro Corp.

Thermal expansion comparisons of the coatings and metal have often been used to determine residual stresses in the coatings. The qualitative residual-stress analysis of Figure 7 shows how enamel A would develop high tensile stress and low residual compressive stress. Enamel B, having a higher softening point and lower expansion coefficient, develops less tensile stress and a much greater residual compressive stress.

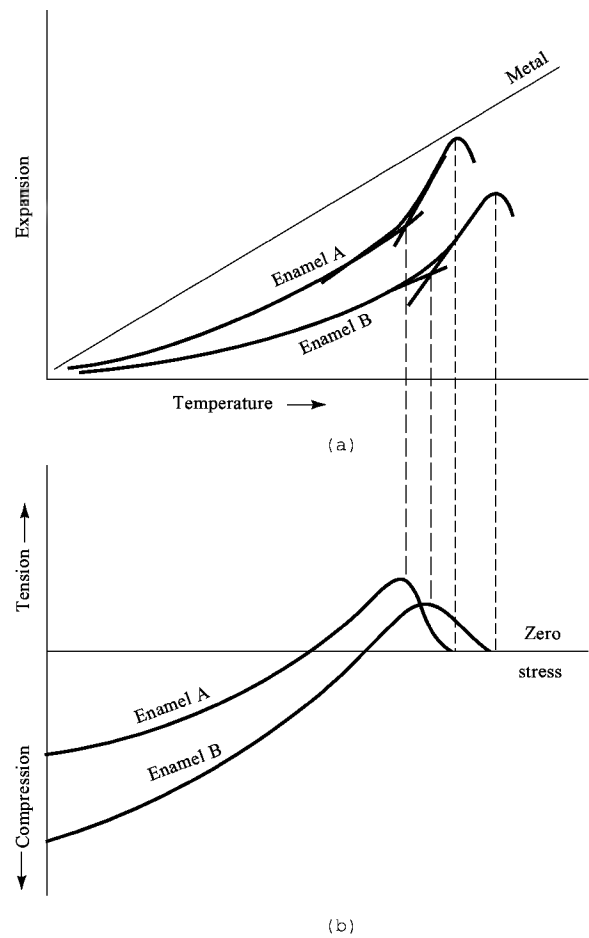


Fig. 7. (a) Relative thermal expansion of porcelain enamel on sheet steel. (b) Stress development in the enamel layer.

Residual-stress analysis must take into account the cooling rates, the viscosity characteristics of the glass, the relative thickness of the metal and coating, and the modulus of elasticity of the coating and metal throughout the temperature range in which stress is developed. At high temperatures, glass behaves as a viscous material, exhibiting viscous-elastic and then primarily elastic properties as the temperature decreases.

Thermal expansion values can be calculated from measurements of thermal deflection of enamel-metal composites. The cubical thermal expansion coefficient in the temperature range of 0–300°C can also be calculated using the additive formula:

$$P = AX_A + BX_B + CX_C + \dots$$

where *P* is the property, such as cubical thermal expansion coefficient; *A*, *B*, *C*, ... are the property factors for each ingredient in the composition; and *X_A*, *X_B*, *X_C*, ... are ingredient concentrations in wt %. Factors for calculating the cubical thermal expansion coefficient of glasses have been determined and are listed in Table 4 (28). Glass compositions high in SiO₂ content have low thermal expansion, whereas additional alkalis and alkaline-earth materials raise the thermal expansion coefficient. Frequently, thermal expansion is listed as linear thermal expansion, which is the cubical thermal expansion divided by three.

Table 4. Enamel Oxide Coefficient of Expansion Factors^{a, b}

Oxide	Factor	Oxide	Factor	Oxide	Factor	Oxide	Factor
SiO ₂	0.8	CaO	5.0	SnO ₂	2.0	Cr ₂ O ₃	5.1
Al ₂ O ₃	5.0	MgO	0.1	TiO ₂	4.1	CoO	4.4
B ₂ O ₃	0.1	BaO	3.0	ZrO ₂	2.1	CuO	2.2
Na ₂ O	10.0	As ₂ O ₃	2.0	Na ₃ AlF ₃	7.4	Fe ₂ O ₃	4.0
K ₂ O	8.5	P ₂ O ₅	2.0	AlF ₃	4.4	NiO	4.0
PbO	4.2	Sb ₂ O ₃	3.6	CaF ₂	2.5	MnO	2.2
ZnO	2.1			NaF	7.4		

^a Courtesy of Garrard Press (15).
^b Summation of (factors × wt%) × 10⁻⁷ = cubical expansion coefficient in vol. (vol.°C).

Residual compression in the coating is desirable because glass, as well as other ceramic materials, is much stronger in compression (about 2070 MPa (300,000 psi)) than in tension (about 69 MPa (10,000 psi)). If residual compression in the enamel layer is too great, the coating may fracture or spall where the radius of curvature of the article is too small. Failure of the coating occurs because the tensile-stress limit has been exceeded. High residual compressive stresses in the coating increase the tensile load-bearing ability by the amount of the residual compressive stress that must be overcome before tension can be induced. If the tensile stress developed during the cooling or reheating of the enamel is too high, the coating fractures. Crazeing results when the damage occurs during cooling; a fine crack pattern called hairlining may be produced during reheating.

The calculated linear average thermal expansion coefficient in cm. (cm.°C) for the temperature range of 25–300°C for common materials in the porcelain enamel system is steel, 11.7 × 10⁻⁶; ground coat, 10–12.5 × 10⁻⁶; cover coat, 8–11 × 10⁻⁶; aluminum, 23.5 × 10⁻⁶; and cast iron, 10.5 × 10⁻⁶.

Composite Modulus of Elasticity. The modulus of elasticity of the enamel glass-steel composite system has been shown to lie between the modulus of the glass and that of the metal (29). The composite modulus can be calculated by

$$E_c = (E_m - E_e) Q^3 + E_e$$

where *E_c*, *E_m*, and *E_e* are the modulus of elasticity of the composite, the metal, and the enamel, respectively; and *Q* = thickness of the metal divided by total thickness of the composite.

Residual Compressive Stress. Residual compressive stress in commercial ground coat enamels varies with enamel thickness:

Ratio of enamel thickness to metal thickness	Compressive stress, MPa (psi)
0.8	69 (10,000)
0.6	110 (16,000)
0.4	138 (20,000)
0.2	221 (32,000)

Thinner coatings have higher compressive stresses, other factors being equal. Higher residual compressive stress in the coating also can be obtained by using enamel glass having a lower thermal expansion coefficient, or a metal having a higher expansion coefficient or a higher modulus of elasticity.

Maximum Strain. Strain in enamels that leads to failure is on the order of 0.002–0.003 cm/cm. Thinner enamels having higher residual compressive stresses are more flexible and can be strained to a greater degree.

Some other physical properties of enamel glass are density, from 2.5–3.5 g/mL; Mohs' hardness, 5–6; tensile strength, 34–103 MPa (4,900–15,000 psi); compressive strength, 1380–2760 MPa (2–4 × 10⁵ psi); modulus of elasticity, 55–83 GPa (8–12 × 10⁶ psi); and dielectric constant, 5–10.

Appearance and Color. Porcelain enamel allows the designer great variety with regard to color, texture, and aesthetic appeal. The enamels exhibit exceptional color stability whether used in domestic interiors or architectural exteriors. Colors should be grouped according to the type and the usage of enamel. Categories are (1) ground coats, which can generally be used by themselves as a finish coat or a base coat; (2) cover coats that are self-opacified; and (3) clear and semi-opaque enamels, used for developing a wide range of colors. Transition-metal oxides used principally to develop a bond between the glass and the substrate metal also produce specific colors. Those range in shades of grays from blue to yellow and red to green. Combinations of each are also possible. Cobalt oxide is the principal colorant for blue, nickel oxide, and iron oxide, Fe₂O₃, for green, manganese dioxide for red, and nickel oxide with other oxides for yellow.

Cover coats such as self-opacified titanium enamels derive their color (qv) from titanium dioxide crystals nucleated in the glass during firing of the coating. This method of opacification is of significant importance to the enamel industry. The titanium-opacified cover coats which exhibit a very high (80–85%) reflectance or Rd (reflectance which is a measure of brightness) are the most common type of white enamel in general use. Other types of opacified enamels are zirconia-opacified enamels (Rd 75–80%) and antimony-opacified enamels (Rd 65–75%). The difference in the refractive indexes between the base glass (1.50–1.55) and the crystal determine the covering power. The greater the covering power, the thinner the coating needed. The most effective opacifiers for enamels (30) are as follows:

Opacifier	Index of refraction
NaF	1.336
CaF ₂	1.434
Sb ₂ O ₃	2.087–2.35
SnO ₂	1.997–2.093
ZrO ₂	2.13–2.20
CeO ₂	2.33
TiO ₂ (anatase)	2.493–2.554
TiO ₂ (rutile)	2.616–2.903

As the level of opacifier is lowered in a glass system, the level of opacification drops. This yields semi-opaque glasses which allow the development of soft pastel colors. Strong colors result from clear glasses. The clarity allows the use of ceramic pigments (qv) (mixed-metal oxides) for the development of a wide variety of colors in almost any hue, saturation, and brightness.

Textures in enamels are developed by the use of semicrystalline glasses or by the addition of refractory materials such as quartz, alumina, zirconia and zircon, feldspar, various clays, and titania. More refractory glasses are also added to impart a lower gloss and a texture. Combinations of glasses such as ground coats and small amounts of self-opacifying cover coat are used to produce speckled finishes such as those found in some oven interiors and cookware utensils. A typical mill batch of colored enamels contains 100 parts by weight frit (clear alkali borosilicate-type glass); 4 parts clay for producing a stable suspension; j part bentonite for supporting suspension; j part sodium nitrite (to disperse flocs); 3 parts oxide pigment colorant; and 45 parts water.

Colored enamels are produced by tinting the titania enamels during the smelting operation through the addition of colorant oxides. Control of enamel color is most popularly done using computer-controlled color measuring systems consisting of spectrophotometric detectors (380–720 nm) and hardware to measure, integrate, evaluate, report, and store the color data. The most common systems in use are the Commission Internationale de l'Eclairage (CIE) *L*a*b** and Rd,*a,b* color scales and various illuminates (31).

Decorating. The decoration of enamels is primarily done by silk screening. Other methods include indirect printing such as use of decals (decalcomania), and indirect lithographic, thermoplastic, and total transfer printing. Photographic printing is also done as a specialty. The range of colors available is nearly unlimited. Textures can also be developed in the decorating materials (32). Older methods of decorating include stenciling, brushing of a second layer to expose an underlying coating, stamping with a rubber transfer device, and graining and marbelizing (33).

Screen printing is used in the highly specialized applications of graphics and circuit designs for electronic devices on a porcelain enameled surface (see INTEGRATED CIRCUITS). Graphic design reproductions may entail between one and ten firings of overlapping prints to achieve a desired result. Halftone printing allows the development of a wider range of colors and effects than is usually achieved using solid colors. Screen printing of rounded surfaces has been developed for special application needs.

The paste used for decorating is typically a finely ground glass softer than the substrate glass, which is mixed with varying color pigments and suspended in an oil medium. The suspending medium is normally a pine oil commonly blended with resins, film formers, and surfactants. A hot oil can be used if the application screen is heated. The hot oil is composed of a variety of fatty alcohols and acrylic resins. Ultraviolet-sensitive vehicles, set by exposure to uv radiation, are composed of photosensitive multifunctional acrylates and resins which may be photoreactive. Some aqueous spray mediums are in use as are some roll-coating formulations which are pine oil based (34).

Microstructure and Thickness. The microstructure of enamels is of importance to understanding and thus being able to control the macroproperties of the enamel. The microstructure is also related to the thickness of the enamel and its firing history. Porcelain enamels typically have a bubble structure which is a result of gas evolution during firing. The gas is mainly a result of the decomposition of organic and inorganic materials. Other sources of gas are the moisture in wet enamel systems and carbon oxide gases from the carbon in the steel. The organics are mostly associated with the clays that are used for suspension of the glass slurry or slip. The inorganic carbonates produce CO₂ bubbles. These bubbles help to inhibit crack propagation in the glass and are reservoirs for the hydrogen gas which is gradually liberated from steels after firing. Figure 8 illustrates typical polished cross sections of various enamel systems.

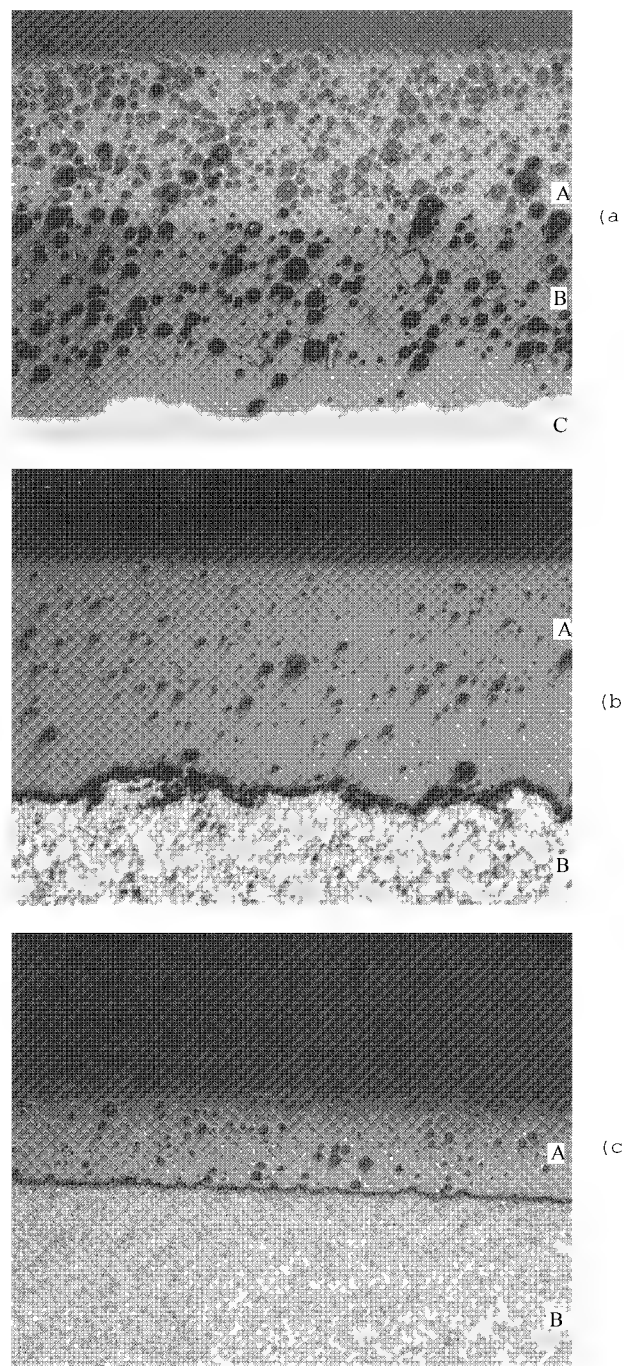


Fig. 8. Enamel cross sections, showing (a) two-coat-two-fire ground coat and cover coat on steel, 30° angle, reflected, polarized light, 100 × magnification where A indicates the 125-μm titania cover coat; B, the 125-μm cobalt ground coat; and C, the steel substrate. (b) Dry-process white cast-iron enamel on gray casting, enamel thickness 755 μm, reflected, polarized light, 30° angle, 25 × magnification where A is the titania cast-iron cover coat and B, the cast-iron substrate. (c) A 79-μm lead-bearing almond aluminum enamel, A, over aluminum alloy 3003, B; reflected light, 30° angle, 100 × magnification.

Courtesy of Ferro Corp.

The thickness of the enamel layer varies with the type of use and metal. However, typical thicknesses are from 75 to 150 μm (3 to 6 mils) for sheet steel, 175 to 359 μm (7 to 14 mils) for hot-rolled steel, 100 to 125 μm (4 to 5 mils) for each coat of wet process cast iron (760 to 788°C fire), 15 to 25 μm (0.5 to 1 mil) for dry process cast-iron base coats and 750 μm (average) to nearly 2250 μm (30 mils to 90 mils) for the dusted cover coats (898 to 955°C fires), and 25 to 50 μm (1 to 2 mils) for aluminum alloys. Stainless steel and copper may have enamel coatings from 40 μm to 175 μm (1.5 to 7 mils) thick.

The study of cross sections is extremely valuable as a way of tracking defects and determining sources of contaminant once the enameled article has been produced. Optical microscopy (qv) of about 10× to 1000× is useful for interpretation of metal grain structures, gas defect sources, and the morphology of the enamel metal interface. Electron micrography having up to hundreds of thousands magnification and the capability of energy dispersive analysis (edax) can be used to identify compositions of minute particles in a polished cross section.

Enamel Testing

Standards. The development of standards for porcelain enamel coatings is shared by several national and international organizations. The American Society for Testing and Materials (ASTM), the Porcelain Enamel Institute (PEI), and the American National Standards Institute (ANSI), as well as the Association of Home Appliance Manufacturers (AHAM) are active in developing, collecting, and disseminating information to interested organizations. Cooperation with the International Standards Organization (ISO) is also fostering the development of internationally unified standards for vitreous enamel coatings.

Abrasion Resistance. Porcelain enamel is the most scratch resistant and hardest of commercial coatings (see **HARDNESS**). This property is used to distinguish between porcelain enamel and organic enamel or painted coatings. The rate of abrasive wear in surface abrasion increases with time, and the subsurface abrasion which follows exhibits a higher, but constant rate of wear. Abrasion resistance can be evaluated by loss of gloss or weight (35).

Adherence. Tests for adherence of porcelain enamels include falling weight tests (36) and hydraulic deformation tests (37,38). The hydraulic deformation test is used for heavy-gauge enameled panels and the drop test for light-gauge metals (18–22 ga). The drop weight test uses an impact energy of 0.9 cm·kg (80 in·lb), with a 1.27 cm (0.50 in.) hemispherical indenter and a 1.9 cm (0.75 in.) diameter bottomless die. Evaluation of the adherence is typically done by a visual rating which assesses the amount of glass retained in the impact area, ranging from 1 (excellent adherence indicated by a nearly solid layer of glass shards) to 5 (poor adherence having little or no glass shards adhering to the metal).

Impact Resistance. Tests for impact resistance of porcelain enamels include falling weight tests such as a free-falling ball or a pendulum

striking a rigidly held specimen. In such tests successively higher heights of fall are used until a visibly noticeable failure occurs. Factors affecting the impact resistance of porcelain enamels include: (1) lower modulus of elasticity of the metal contributes to higher impact resistance; (2) the larger radius of curvature of the article, the greater the impact resistance; (3) the thicker the metal and enamel, the greater the resistance to impact; and (4) the physical structure of the enamel, eg, excessive bubble content, large crystalline inclusions, and other discontinuities, often contribute to lower impact resistance. Weak bonding of the enamel to the metal also contributes to lower impact resistance.

Thermal Shock Resistance. Resistance of porcelain enamels to failure by thermal shock was developed for enamels for cooking utensils and other items subjected to high temperatures in service. Thermal shock is experienced by a heated enamel article on quenching with cold water. Thermal shock tests involve repeated cycles of heating and quenching with water; heating for each successive cycle is carried out at progressively higher temperatures. A visible fracturing, as evidenced by spalling, constitutes a failure.

Thermal shock failures using water result from the water vapor entering the enamel layer through small, submicroscopic cracks formed at the instant of shock. The water condenses in the cracks and in the bubbles of the enamel traversed by the cracks. On subsequent heating, the vapor from the entrapped water expands to cause spalling of the enamel layer. Other quenchant liquids, such as toluene, oils, and other organic liquids, also cause fine, almost invisible cracks, but thermal shock failures do not result with these quenchants on subsequent heating (39).

Thermal shock resistance is a direct function of enamel thickness. The greater the residual compressive stress in the porcelain enamel, the greater is the resistance to thermal shock failure. Thin coatings, such as one-coat enamels or the two-coat enamels having a low expansion titania cover coat, provide excellent thermal, shock resistance.

Resistance to Chemical Attack. The resistance to alkali and acid attack is evaluated on the basis of loss in weight, loss in gloss, or reduced cleanability of the surface.

A spot test employing 10° citric acid is used at room temperature for acid resistance of glossy, light-colored enamels. In the spot test, loss of gloss and cleanability are determined in a qualitative manner (40). Resistance to boiling 6° citric acid is determined by loss in weight (41). Lower alkali content of the glass yields higher acid resistance. Acid-resistant enamels for chemical service are compositions high in SiO₂ and TiO₂ content. Alkali resistance is improved with increasing Al₂O₃ and ZrO₂ content. Resistance of enamels to water attack is also introduced in coatings for domestic water heaters.

Other Properties and Tests. The other physical and chemical properties of porcelain enamels can be evaluated as shown below:

ASTM test number	Property
C614-74 (1989)	alkali resistance
C538-83	color retention of red, orange, and yellow enamels
C743-87	continuity of porcelain enamel coatings
C374-70 (1982)	fusion flow of porcelain enamel frits (flow-button methods)
C346-87	gloss of ceramic materials, 45° specular
C372-81	linear thermal expansion by the dilatometer method
C285-88	sieve analysis of wet-milled and dry-milled porcelain enamel
C703-72 (1983)	spalling resistance (enameled aluminum)
C385-58 (1983)	thermal shock resistance of porcelain enameled utensils
C448-88	abrasion resistance

Enamel Defects. Characterization of defects in porcelain enamel surfaces frequently requires detailed examination via microscopy to determine the sources of the defects. Defects are divided into processing and material defects. The greatest number of defects result from processing: blisters, pinholes, black specks, dimples, tool marks, and chipping. Defects often occur from unobserved sources at almost every stage of the enameling process, but they are not recognizable until the ware is fired. Conscientious process control helps to minimize the incidents of unacceptable finishes.

Economic Aspects

The porcelain enameling market has been gradually declining since 1974 in the United States in terms of tonnage. Self-smelters represent about half of the total production and specialty producers the remainder. The total U.S. market in 1990 was about 90,000 metric tons whereas the average selling price of porcelain enamel frit was about \$2.00–2.50 kg. Substitution of less expensive, and often less durable, organic coatings and plastics has been mainly responsible for the decline in the usage of porcelain enamel.

Shipment. The shipment of porcelain enamel is typically in 45-kg (100-lb) three-ply paper bags. More recently, 227-kg (500-lb) multi-ply cardboard barrels and bulk bags made of high strength woven polymers have been used. Porcelain enamel frit, powder, and premixed and premilled mill additions are shipped in these flexible containers.

Labeling. The typical labeling classification for frit may be followed by precautionary labeling for formulations containing lead oxide, free silica, or cadmium oxide. Special labeling for shipment to specific localities may also be necessary to meet local and state requirements.

BIBLIOGRAPHY

"Enamels, Porcelain or Vitreous" in *ECT* 1st ed., Vol. 5, pp. 718–735, by R. M. King, The Ohio State University; in *ECT* 2nd ed., Vol. 8, pp. 155–173, by A. L. Friedberg, University of Illinois, in *ECT* 3rd ed., Vol. 9, pp. 1–20, by A. L. Friedberg, University of Illinois.

1. *1992 Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, Pa., 1992, p. 02.05.
2. M. A. Baker, *Proc. Porcelain Enamel Inst. Forum* **34**, 186–198 (1972).
3. J. E. Waggener, *Proc. Porcelain Enamel Inst. Forum* **52**, 861–863 (1990).
4. J. Wratil, *Vitreous Enamels*, Borax Holdings Limited, London, 1984, p. 13.
5. H. F. Evele and W. D. Faust, *Proc. Porcelain Enamel Inst. Forum* **47**, 643–650 (1985).
6. R. G. Rion and E. Smithberger, *Proc. Porcelain Enamel Inst. Forum* **38**, 69–74 (1976).
7. L. N. Smith, *Proc. Porcelain Enamel Inst. Forum* **52**, 661–663 (1990).
8. P. Vadeboncoeur and R. Hirt, *Proc. Porcelain Enamel Inst. Forum* **36**, 75–79 (1974).
9. F. W. Nelson and L. N. Beer, *Proc. Porcelain Enamel Inst. Forum* **33**, 111–115 (1971).
10. G. R. Prinzbach, *Proc. Porcelain Enamel Inst. Forum* **53**, (1991).
11. *Pretreatment of Alloys for Porcelain Enameling Aluminum*, PEI Bulletin P-403(70) and *Enamel Preparation, Application, and Firing for Porcelain Enameling Aluminum*, PEI Bulletin P-404(70), Porcelain Enamel Institute, Washington, D.C., 1970.
12. *1990 Book of ASTM Standards*, Publication A424-89, American Society for Testing and Materials, Philadelphia, Pa., 1990, pp. 168–169.
13. Ref. 12, p. 169.
14. R. W. Nuske, T. A. Bloom, and T. L. Schiller, *Proc. Porcelain Enamel Inst. Forum* **52**, 836–846 (1990).
15. A. I. Andrews, *Porcelain Enamels*, 2nd ed., Garrard Press, Champaign, Ill., 1961, p. 111.
16. *Alloy Design and Fabrication Considerations for Porcelain Enameling Aluminum*, Bulletin P-402(69), Porcelain Enamel Institute, Washington, D.C., 1969.
17. Ref. 12, pp. 136–137.
18. Ref. 12, p. 210.
19. *Pretreatment of Alloys for Porcelain Enameling Aluminum*, Porcelain Enamel Institute Bulletin P-403(70), 1970.

20.

L. C. Kolar and co-workers, *Proc. Porcelain Enamel Inst. Forum* **49**, 420–424 (1987).

21.

M. Bozin, *Chemical and Physical Determinations for Porcelain Enamel Plants*, Ferro Corp., Cleveland, Ohio, 1951, p. 12.

22.

Ref. 20, p. 422.

23.

B. W. King, H. P. Tripp, and W. H. Duckworth, *J. Am. Ceram. Soc.* **42**(11), 504–525 (1959).

24.

A. L. Gugeler, *Proc. Porcelain Enamel Inst. Forum* **31**, 37–45 (1969).

25.

Ref. 12, pp. 96–97.

26.

1992 Book of ASTM Standards, Publications C372-88 and C824-91, American Society for Testing and Materials, Philadelphia, Pa., 1992, p. 15.02.

27.

E. E. Bryant, *Porcelain Enameling Operations*, Enamelist Publishing Co., Cleveland, Ohio, 1964.

28.

Ref. 12, p. 53.

29.

P. S. Wolford and G. E. Selby, *J. Am. Ceram. Soc.* **29**(6), 162 (1946).

30.

Ref. 4, p. 39.

31.

A. R. Robertson, *Color Res. Appl.* **15**(3), 167–170 (June 1990).

32.

J. Tunncliffe and M. Rose, *Vitreous Enameller* **40**(1), 1–5 (Spring 1989).

33.

Ref. 12, pp. 408–413.

34.

J. VanNess, Ferro Corp., private communications, Nov. 1991.

35.

Ref. 1, Publication C448-88, 1992.

36.

Adherence to Porcelain Enamel Cover Coats Direct to Steel, PEI Bulletin T-29(72), Porcelain Enamel Institute, Washington, D.C., 1972.

37.

Adherence of Porcelain Enamel and Ceramic Coatings to Sheet Metal, PEI Bulletin T-17(72), Porcelain Enamel Institute, Washington, D.C., 1972.

38.

Ref. 12, Publication C313, 1990.

39.

Ref. 1, Publication C385-58, 1992.

40.

Ref. 1, Publication C282-89, 1992.

41.

Ref. 1, Publication C283-54, 1992.

General References

United States

A. I. Andrews, *Porcelain Enamels*, 2nd ed., Garrard Press, Champaign, Ill., 1961.

J. F. Wright, C. G. Bergeron, and J. C. Oliver, "Porcelain Enamel," in S. J. Schneider, Jr., vol. chrmn., *Engineering Materials Handbook*, Vol. 4, ASM International, Materials Park, Ohio, 1991, pp. 937–942.

J. C. Oliver, "Porcelain Enamels-Review and Outlook," *Am. Ceram. Soc. Bull.* **62**(5), 562–563 (1983).

D. Balmer, *Proc. Porcelain Enamel Inst. Forum* **50**, 492–497 (1988).

L. Gypen, *Proc. Porcelain Enamel Inst. Forum* **52**, 885–897 (1990).

E. W. Hughes and D. F. Thomas, *Proc. Porcelain Enamel Inst. Forum* **49**, 391–393 (1987).

A. Bets and K. O. Helm, *Proc. Porcelain Enamel Inst. Forum* **49**, 380–385 (1987).

S. D. Brown, *Proc. Porcelain Enamel Inst.* **46**, 301–317 (1984).

D. B. Judd and G. Wyszecski, eds., *Color in Business, Science, and Industry*, 2nd ed., John Wiley & Sons, Inc., New York, 1967.

International

J. Watril, *Vitreous Enamels*, Borax Holdings Ltd., London, 1984.

K. A. Maskall and D. W. White, *Vitreous Enamelling: A Guide to Modern Enamelling Practice*, Pergamon Press, Oxford, UK, 1986.

A. H. Dietzel, *Emaillierung: Wissenschaftliche Grundlagen und Grundzüge der Technologie*, Springer-Vedag, Berlin, 1981.

William D. Faust
Ferro Corporation

ENCAPSULATION

. See EMBEDDING; MICROENCAPSULATION.

ENDORPHINS

. See HORMONES, BRAIN PEPTIDES; HYPNOTICS, SEDATIVES, AND ANTICONVULSANTS; NEUROREGULATORS; OPIOIDS, ENDOGENOUS.

ENERGY MANAGEMENT

In the chemical industry, which is inherently energy intensive, energy costs, including feedstock, average approximately 8% of the value added. For large-volume chemicals these costs represent a much higher fraction. For example, for nitrogen-based fertilizer the energy costs are approximately 70% of the value added (see FERTILIZERS).

Energy management includes energy conservation, but also encompasses utility system reliability; the intermesh of process design with utility systems; purchasing, including plant location for minimum energy cost; environmental impacts of energy use; tracking energy performance; and the optimization of energy against capital in equipment selection (see also ECONOMIC EVALUATION; POWER GENERATION; PROCESS ENERGY CONSERVATION).

Energy and the Chemical Industry

Table 1 is an estimate of energy usage by United States industry for 1988 (1). The chemical industry used 21% of the energy consumed by the U.S. industrial sector, and the other three related process industries, paper (qv), petroleum (qv), and primary metals, combined for an additional 50% of the industrial consumption.

Table 1. Energy Usage by United States Industry in 1988^a

Energy source	Consumption, EJ ^{b,c}				
	Chemicals	Petroleum refining	Paper	Primary metals	Other
electricity					
direct use	0.44	0.11	0.20	0.54	1.24
loss ^d	0.88	0.22	0.40	1.08	2.48
oil	0.15	0.11	0.20	0.06	0.34
gas and LPG	3.11	0.78	0.45	0.79	2.10
coal and coke	0.34		0.34	1.59	0.68
other	0.58	5.65 ^e	1.31 ^f	0.05	0.51
Total	5.50	6.87	2.90	4.11	7.35

^a Values include feedstocks.

^b One EJ = 1 × 10¹⁸ Joules.

^c To convert Joules to Btu, multiply by 0.95 × 10⁻³.

^d Assumes delivered electricity has one-third of the energy content of the fuel used to produce it; ie, 1 Joule of electrical supply to an industrial user represents 3 Joules of fuel consumed.

^e For petroleum, this value represents primarily feedstock used in nonenergy products such as asphalt, as well as some feedstock used for petrochemicals.

^f For paper, this value represents the fuel value of the biomass.

Feedstocks. A separate breakdown between fuels and feedstocks (qv) for the chemical industry (2) shows that the quantity of hydrocarbons (qv) used directly for feedstock is about as great as that used for fuel (see FUELS, SYNTHETIC; GASOLINE AND OTHER MOTOR FUELS). Much of this feedstock is oxidized accompanied by the release of heat, and in many processes, by-product energy from feedstock oxidation dominates purchased fuel and electricity. A classic example is the manufacture of nitric acid (qv) [7697-37-2], HNO₃. Ammonia (qv) [7664-41-7], NH₃, is burned in air on a catalyst at a pressure around 1 MPa (10 atm) and a temperature above 900°C (3). As shown in Figure 1, the process is built around the heat and power recovery from this reaction, including the integration of the power recovery turbine and the air compressor. Another example is the reaction of propylene (qv) [9003-07-0], C₃H₆, and ammonia to make acrylonitrile (qv) [107-13-1], C₃H_{3.5}N. Here, two-thirds of the hydrogen is combusted just to satisfy stoichiometry. In addition, CO and CO₂ formation consumes about 15% of the feed propylene.

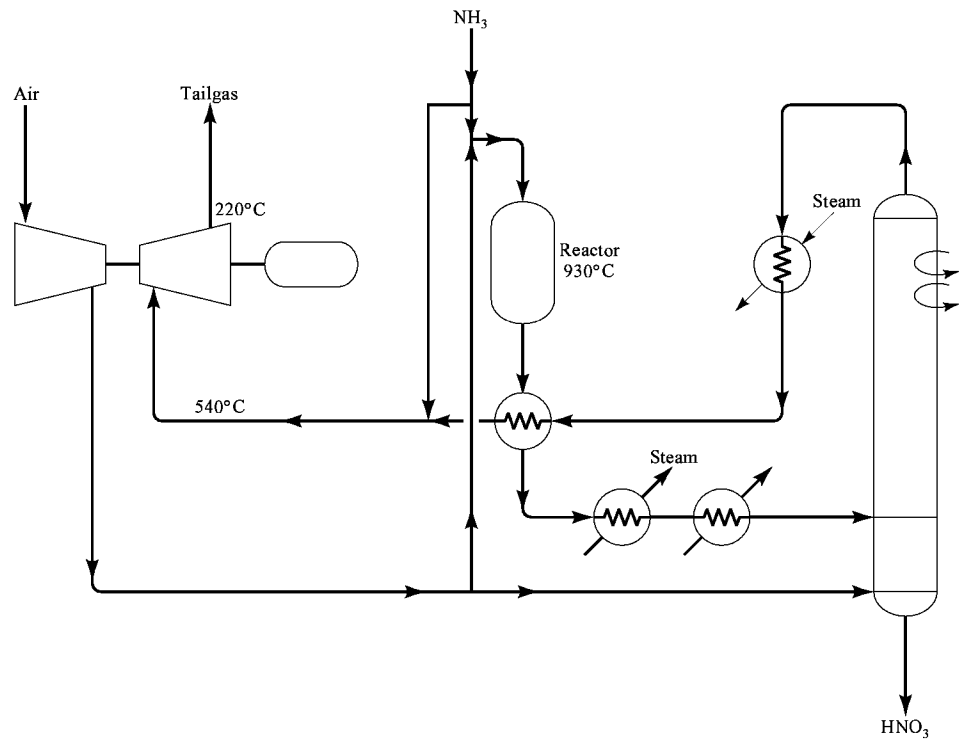


Fig. 1. Schematic of nitric acid from ammonia showing integration of reactor heat recovery, power recovery from tailgas, and air compression (3).

Fuels. Two-thirds of the fuel used by the United States chemical industry in 1988 was natural gas [8006-14-2], which is clean and easy to combust (see GAS, NATURAL). Although relatively inexpensive at the wellhead, natural gas is costly to transport. Hence the chemical industry is concentrated in regions where natural gas is produced, keeping the average price paid by the U.S. chemical industry for natural gas in 1988 to only 80% of the average U.S. industrial price (1). Similarly the movement of chemical commodity production to the Middle East is driven by the desire to obtain low cost natural gas.

Waste Fuel Utilization. It is always preferable to minimize or upgrade by-products for sale as chemicals, however when this is not feasible the fuel value can still be recovered. Increased combustion of by-product gases and liquids was one of the principal components in the improvement in energy efficiency that occurred in the industry in the 1980s. An example of waste fuel utilization is the incineration of the off-gases from an acrylonitrile reactor followed by generation of high pressure steam, shown in Figure 2.

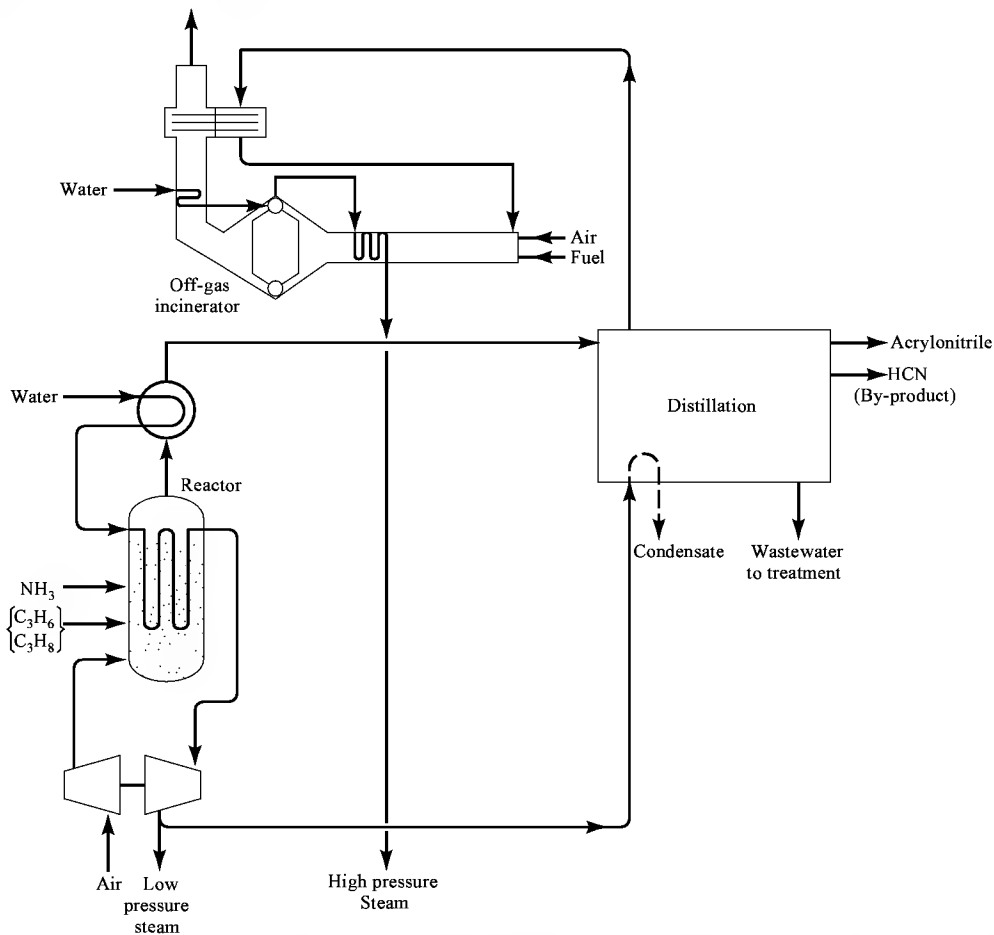


Fig. 2. Acrylonitrile process showing integration of waste heat from reactor and off-gas incinerator. The reaction in the reactor is $\text{C}_3\text{H}_6 + \text{NH}_3 + 1.5 \text{O}_2 \rightarrow \text{C}_3\text{H}_3\text{N} + 3 \text{H}_2\text{O} + \text{heat}$; in the off-gas incinerator, $\text{C}_3\text{H}_8 + 5 \text{O}_2 \rightarrow 3 \text{CO}_2 + 4 \text{H}_2\text{O} + \text{heat}$.

Electricity. Electricity, including the losses associated with production, represents 24% of the total energy used by the chemical industry (Table 1). On a cost basis, electricity represents a higher share at 29% of the energy bill including feedstocks. Increases in electrical costs have provided the driving force for increased cogeneration, ie, the recovery of power as a by-product of other process plant operations (see also POWER GENERATION). The historic

cogeneration example is the steam turbine associated with the boiler plant. The relatively high cost of electricity has also led designers to focus on the efficiency of rotating equipment, and has motivated a closer look at how processes can be controlled to reduce power, using such innovations as variable frequency motor drives.

Energy Efficiency Improvements. Energy management is basically a game of economics played with a special set of technical rules. A saving of millions of kilowatt hours or gigajoules is only communicable when converted into dollars, or into a ratio of dollars saved per dollar of incremental investment. The question of where in a process to focus on improvement is often answered by determining the biggest energy cost items in a plant.

Efficiency improvement is driven by two distinct forces: technological progress which is the long-term trend, and cost optimization which is the short-term response to price swings. The baseline, long-term trend has been in the range of 2 to 3% per year improvement. The forecast for the 1990s is 1 to 2% per year, reflecting a more mature industry. A 1 to 2% per year energy reduction still yields a large saving over time.

The U.S. chemical industry achieved an annual reduction of 4.2% in energy input per unit of output for the period 1975–1985 (2). This higher reduction resulted from cost optimization, the tradeoff of increased capital for reduced energy use, that was driven by energy prices (4). In contrast, from 1985 to 1990, the energy input per unit of output has been almost flat (2) as a consequence of falling prices. The average price the U.S. chemical industry paid for natural gas fell by one-third between 1985 and 1988 (1,5).

Whereas energy conservation is an important component of cost reduction for the chemical industry, conservation is rarely the only driving force for technological change. Much of the increased energy efficiency comes as a by-product of changes made for other reasons such as higher quality, increased product yield, lower pollution, increased safety, and lower capital. For example, process energy integration in design, enabled by computer simulation, saves energy as well as capital; substitution of variable speed drives on motors for control valves saves energy as well as improves control; and the use of gas turbines in cogeneration, saves energy as well as capital in the supply of power. One of the roles of energy management is to be sure energy use reduction is considered whenever processes are changed.

The refinement of processes such as occurred for production of low density polyethylene, where a process requiring over 100 MPa (1000 atm) was replaced by one taking place at MPa (20 atm) (see OLEFIN POLYMERS), may continue, but such refinement is expected to be less than in the 1970s and 1980s. The introduction of biotechnology-derived processes is expected to cause a shift to lower temperature and lower pressure processing in the chemical industry, but as of this writing the timing is unclear.

Energy and the Environment. The impact of energy usage on gaseous emissions has emerged as a primary environmental issue, and regulatory action has required emission reductions in NO_x and SO₂ (see AIR POLLUTION; AIR POLLUTION CONTROL METHODS; EXHAUST CONTROL, INDUSTRIAL). Control of NO_x is achieved by limiting the temperature of combustion and limiting excess oxygen. Control of SO₂ either requires changing fuel or flue gas scrubbing. Because the preferred fuel of the chemical industry is already predominantly low sulfur natural gas, the primary impact on the chemical industry of SO₂ regulation is expected to be to raise the price of electricity derived from coal (qv).

Issues related to gases such as CO₂, which contribute to the global greenhouse effect, are also rising in importance (see ATMOSPHERIC MODELS). Energy conservation directly reduces CO₂ emissions. The elimination of fugitive hydrocarbon emissions as a result of improved maintenance procedures is also a tangible step that the industry is taking.

Energy Technology

Energy management requires the merging of such technologies as thermodynamics, process synthesis, heat transfer, combustion chemistry, and mechanical engineering (see also COMBUSTION TECHNOLOGY; HEAT EXCHANGE TECHNOLOGY; SIMULATION AND PROCESS DESIGN; and THERMODYNAMIC PROPERTIES).

Thermodynamics. The first law of thermodynamics, which states that energy can neither be created nor destroyed, dictates that the total energy entering an industrial plant equals the total of all of the energy that exits. Feedstock, fuel, and electricity count equally, and a plant should always be able to close its energy balance to within 10⁰ o. If the energy balance does not close, there probably is a big opportunity for saving.

The second law of thermodynamics focuses on the quality, or value, of energy. The measure of quality is the fraction of a given quantity of energy that can be converted to work. What is valued in energy purchased is the ability to do work. Electricity, for example, can be totally converted to work, whereas only a small fraction of the heat rejected to a cooling tower can make this transition. As a result, electricity is a much more valuable and more costly commodity.

Unlike the conservation guaranteed by the first law, the second law states that every operation involves some loss of work potential, or exergy. The second law is a very powerful tool for process analysis, because this law tells what is theoretically possible, and pinpoints the quantitative loss in work potential at different points in a process.

Typically, the biggest lost that occurs in chemical processes is in the combustion step (6). One-third of the work potential of natural gas is lost when it is burned with unpreheated air. Figure 3 shows a conventional and a second law heat balance. The conventional analysis only points to recovery of heat from the stack as an energy improvement. Second law analysis shows that other losses are much greater.

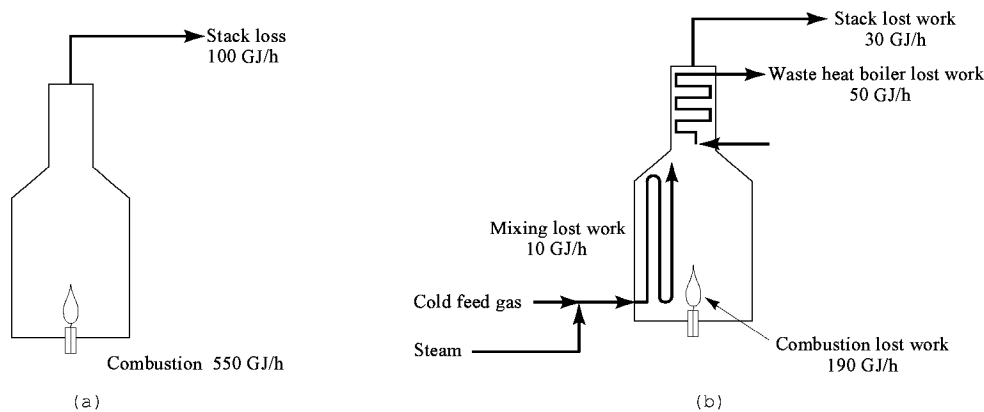


Fig. 3. (a) A conventional heat balance, and (b) a balance employing the second law of thermodynamics. To convert J h to Btu h, multiply by 0.95×10^{-3} .

The second law can also suggest appropriate corrective action. For example, in combustion, preheating the air or firing at high pressure in a gas turbine, as is done for an ethylene (qv) cracking furnace, improves energy efficiency by reducing the lost work of combustion (Fig. 4).

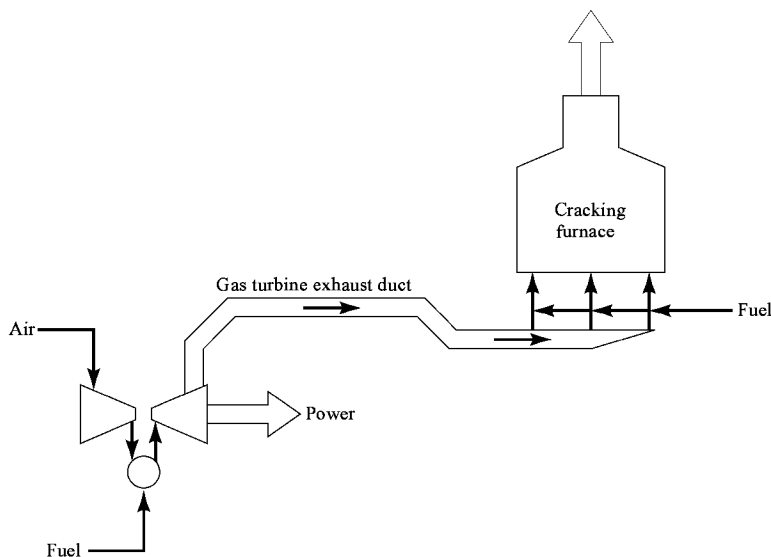


Fig. 4. Use of gas turbine air preheat for ethylene cracking furnace. The gas turbine exhaust duct contains 17% oxygen at 400°C.

Converting Heat to Work. There has been a historic bias in the chemical industry to think of energy use in terms of fuel and steam (qv) systems. A more fundamental approach is to minimize the input of work potential embedded in the fuel and feedstock, as well as work purchased directly as electricity. Steam is really just a medium of exchange, like money in an economy.

Waste heat tends to be very visible. It can be seen directly in steam plumes and easily measured by determining the temperature of the discharges. The loss of work potential or excess use of work is much harder to spot, but often is larger and more easily corrected. Small inefficient turbines, oversized pumps run on minimum bypasses, rewound motors having efficiency far below design, higher than optimum temperature differences that lead to below optimum power recovery, organics discharged in wastewater, and pressure drops taken across control valves are all examples of loss of work potential.

Economics of Energy Levels and Power Recovery.

Cogeneration in a Steam System. The value of energy in a process stream can always be estimated from the theoretical work potential, i.e., the determination of how much power can be obtained by running an ideal cycle between the actual temperature and the rejection temperature. However, in a steam system a more tangible approach is possible, because steam at high pressure can be let down through a turbine for power. The shaft work developed by the turbine is sometimes referred to as by-product power, and the process is referred to as cogeneration.

The by-product power takes only 40% as much energy to produce as on-purpose firing for power only. A comparison of Figure 5b and 5d illustrates this point. By-product power also takes much less equipment, as can be seen in Figure 6. By-product power also avoids the losses and capital associated with use of electricity such as power lines, transformers, and motors. As a result, by-product power is much cheaper than power purchased as electricity.

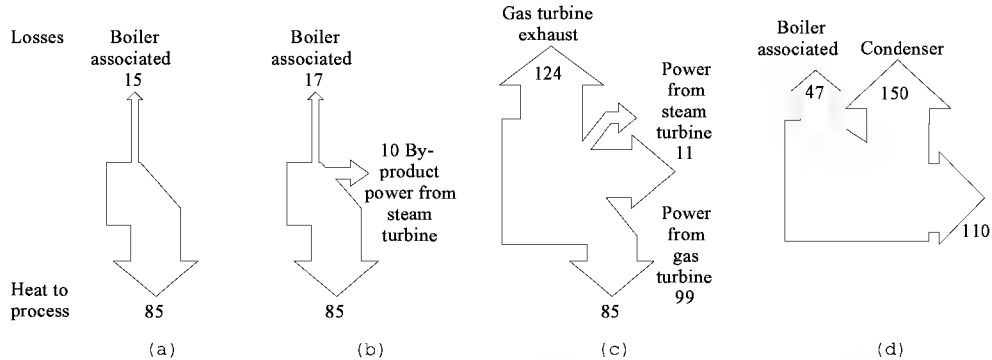


Fig. 5. Relative energy flows showing power generation and heat losses in GJ/h for (a), boiler only; (b), boiler + steam turbine; (c) combined cycle employing gas turbine; and (d), condensing steam for power only. To convert J/h to Btu/h, multiply by 0.95×10^{-3} .

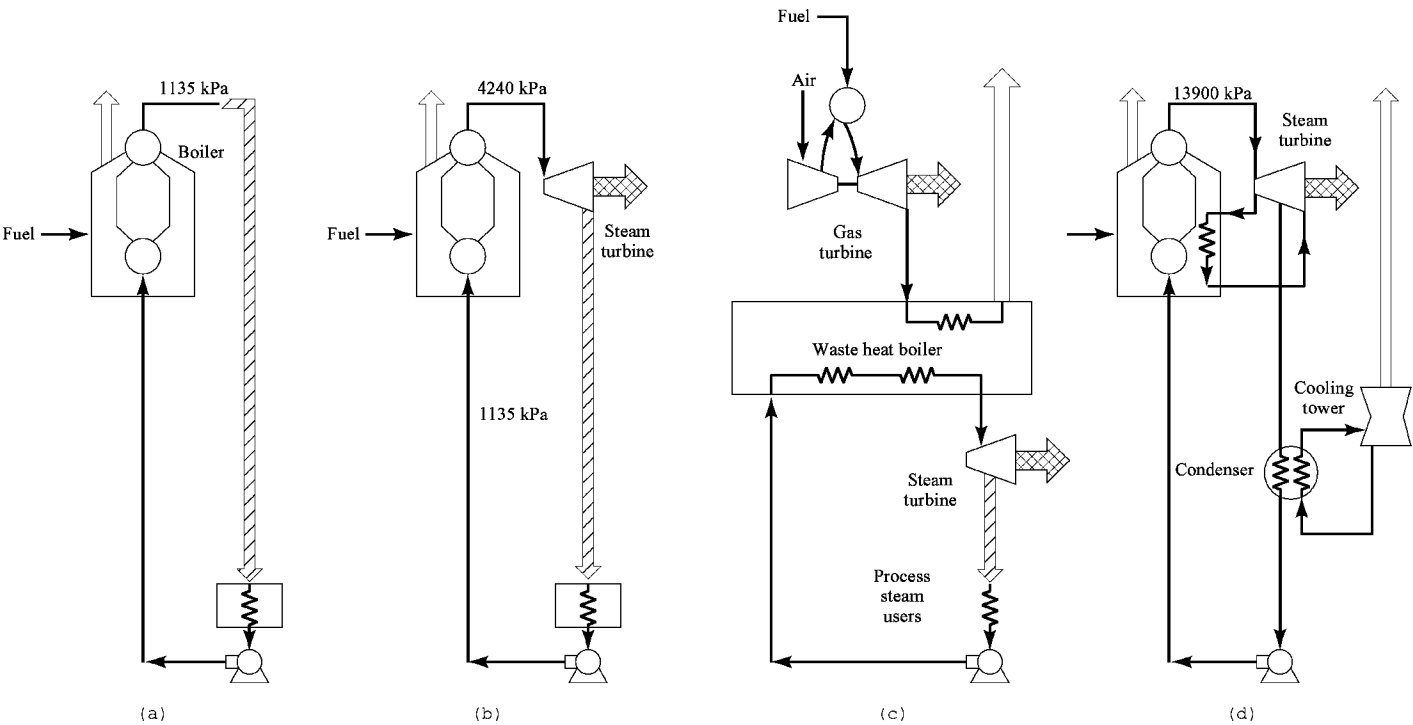


Fig. 6. Schematics showing the principal equipment components for the energy systems shown in Figure 5. (a) through (d) correspond to Figure 5a through 5d, respectively. To convert kPa to psi, multiply by 0.145.

There is, however, only a limited quantity of by-product power available, and for large process operations the demand for power is usually far greater than the simple steam cycle can produce. Many steam system design decisions fall back to the question of how to raise the ratio of by-product power to process heat. One simple approach is to limit the turbines that are used to extract power to large sizes, where high efficiency can be obtained. Another way to raise the power heat ratio is by raising the pressure of the steam system. An increase in pressure from 4.2 to 10.1 MPa (600 to 1500 psi) almost doubles the power associated with a given steam load. (Power heat ratio increases from 0.12 to 0.20). This, however, comes at appreciable capital cost for alloy materials of construction in the boiler, piping, and turbines. It also requires

Parameter	(a)	(b)	(c)	(d)
fuel	100	112	319	307
incremental fuel	0	12	219	307
power	0	10	110	110
power incremental fuel		0.85	0.5	0.36
power heat to process	0	0.12	1.29	

a substantially higher cost treatment system for the boiler feedwater, and mandates a relatively high recovery factor for condensate. By-product power does not give enough power to match the demand for many processes such as ammonia synthesis, and designs have historically incorporated condensing turbines for incremental power with heat rejection to cooling water. A more effective response is use of the gas turbine combined cycle shown by Figures 5c and 6c.

Gas Turbines and the Combined Cycle. The combined cycle first fires fuel into a gas turbine and greatly increases the power extracted per unit of steam produced. As the numbers in Figure 5 show, the simple steam turbine gives by-product power at the lowest incremental energy use. The combined cycle shown in Figure 5c is intermediate in energy per unit of power between Figure 5b and Figure 5d. The big advantage of the gas turbine in cogeneration is that it permits a much higher ratio of power to heat. This ratio, which is routinely >0.8, gets bigger as the unit size of the turbine increases (7). The ratio of power to heat is also larger for aero-derivative systems which are basically jet engines exhausting into power recovery turbines. Because the original design assumed only a gas cycle, ie, no waste heat boiler, the aero-derivative was designed for a higher compression ratio, resulting in a cooler (500°C) gas turbine exhaust. Cooler exhaust means less heat to surrender, and hence a higher power heat ratio. The heavy-duty machines of comparable vintage run discharge temperatures about 100°C higher. The aero-derivatives are also lighter in weight, giving a second advantage for a process operation because the lightweight permits very fast plug-out plug-in maintenance using spare rotating assemblies.

In contrast, the heavy-duty gas turbines, designed with capital cost per kilowatt as a key criterion, cost less for the same power output and come in much larger sizes than the aero-derivatives. This means that very large (>100 MW) installations invariably use the heavy-duty type. Heavy-duty turbines have benefited from such technology developments of the aero-derivatives as increased firing temperatures permitted by use of high temperature alloys (qv) and internal cooling of the blading. These developments have been a factor in continuing increased efficiency and increased output from a given frame size.

Gas turbine cogeneration is inherently relatively low in capital expenditures. The absence of heat exchange surface in the gas turbine part of the cycle provides the basic capital advantage, and standardized equipment, prepackaged as skid mounted components, adds to the capital advantage. When these factors are coupled with low priced natural gas, a situation results in which petrochemical plants have become exporters of cogenerated power to utilities. The gas turbine also has advantages that are firmly rooted in thermodynamics. It utilizes energy directly at a high temperature level, without large driving forces for pressure drop and temperature difference.

Most gas turbine applications in the chemical industry are tied to the steam cycle, but the turbines can be integrated anywhere in the process where there is a large requirement for fired fuel. An example is the use of the heat in the gas turbine exhaust as preheated air for ethylene cracking furnaces as shown in Figure 4 (8).

The combined cycle is also applicable to dedicated power production. When the steam from the waste heat boiler is fed to a condensing turbine, overall conversion efficiencies of fuel to electricity in excess of 50% can be achieved. A few public utility power plants use this cycle, but in general utilities have been slow to convert to gas turbines. Most electricity is generated by the cycle shown in Figures 5d and 6d.

Power Recovery in Other Systems. Steam is by far the biggest opportunity for power recovery from pressure letdown, but others such as tailgas expanders in nitric acid plants (Fig. 1) and on catalytic crackers, also exist. An example of power recovery in liquid systems, is the letdown of the high pressure, rich absorbent used for H₂S CO₂ removal in NH₃ plants. Letdown can occur in a turbine directly coupled to the pump used to boost the lean absorbent back to the absorber pressure.

Heat Recovery, Energy Balances, and Heat-Exchange Networks. The goal of heat recovery is to be sure that energy does the

maximum useful work as it cascades to ambient. An energy balance is a summary of all of the energy sources and all of the energy sinks for a unit operation, a process unit, or an entire manufacturing plant. Table 2 gives an energy balance for a simple propane-fired dryer. The energy balance is almost as important to understanding how a process works as the material balance. The energy balance is the basic tool for analyzing an operation for energy conservation opportunities. When incorporated into a computer program, the energy balance becomes the base for a model of the process. Operational changes, system configuration, and equipment alterations can be evaluated via the model.

Table 2. Product Dryer Analysis Heat Balance

Material	Mass, kg h	Energy, MJ h ^a
<i>Inputs</i>		
fuel, C ₃ H ₈	130	6,553
air		
combustion	6,817	106
secondary	14,846	232
in-leakage	4,289	67
water with product	1,731	354
dry product solids	4,478	249
Totals in	32,291	7,561
<i>Outputs</i>		
water vapor		
from product	1,445	3,808
from combustion of H ₂	212	560
air and combustion products	25,870	1,933
dry product solids	4,478	458
water with product out	286	146
heat losses		656
Totals out	32,291	7,561

^a To convert J h to Btu h, multiply by 0.95 × 10⁻³.

The heat exchanger network analysis, sometimes called pinch technology, is a special kind of model that has been developed into a sophisticated way of attacking heat recovery problems (see HEAT EXCHANGE TECHNOLOGY, NETWORK SYNTHESIS). This type of analysis defines the optimum interchange between all the heat sinks and heat sources. It is similar to the concept used by furnace designers for many years for matching various heat sinks against the flue gas source. Using a temperature vs enthalpy plot, the analysis can be done manually or via computer. It can even be set up to adjust distillation sequence or operating pressures to improve energy efficiency and minimize capital (9).

Waste-Heat Boilers. In most chemical process plants, the steam system is the integrating energy system. Recovering waste heat by generating steam makes the heat usable in any part of the plant served by the steam system. Many waste-heat recovery boilers are unique and adapted to fit a particular process. There is a long history of process waste-heat boiler failure resulting from inadequate attention to detail in design and the failure to maintain water quality (10). The high heat-transfer coefficients of boiling water are dependent upon clean surfaces. Designers should match the hardware as closely as possible to demonstrated designs, and operators should be sure that water treatment is monitored. Incinerators (qv) and gas turbines also involve heat recovery boilers. Here, a number of fairly standard designs have evolved (11).

Product-to-Feed Heat Interchange. Heat exchange is commonly used to cool the product of a thermal process by preheating the feed to that process, thus providing a natural stabilizing, feed forward type of process integration. Product-to-feed interchange is common on reactors as well as distillation (qv) trains.

Combustion Air Preheat. Flue gas to air exchange, a type of product-to-feed heat exchange, is extremely important because of the large loss associated with the combustion of unpreheated air. This exchange process has generated fairly unique types of hardware such as the Ljungstrom or rotary wheel regenerator shown in Figure 7a; the brick checkerwork regenerators used in metallurgical furnaces, hot oil, or hot water belts (also called "liquid runarounds") (Fig. 7b); and heat pipes (Fig. 8). Liquid runaround systems make it practical to use finned surface on both gas exchange surfaces. These are particularly useful for retrofit because of the ability to move the heat to physically separated units.

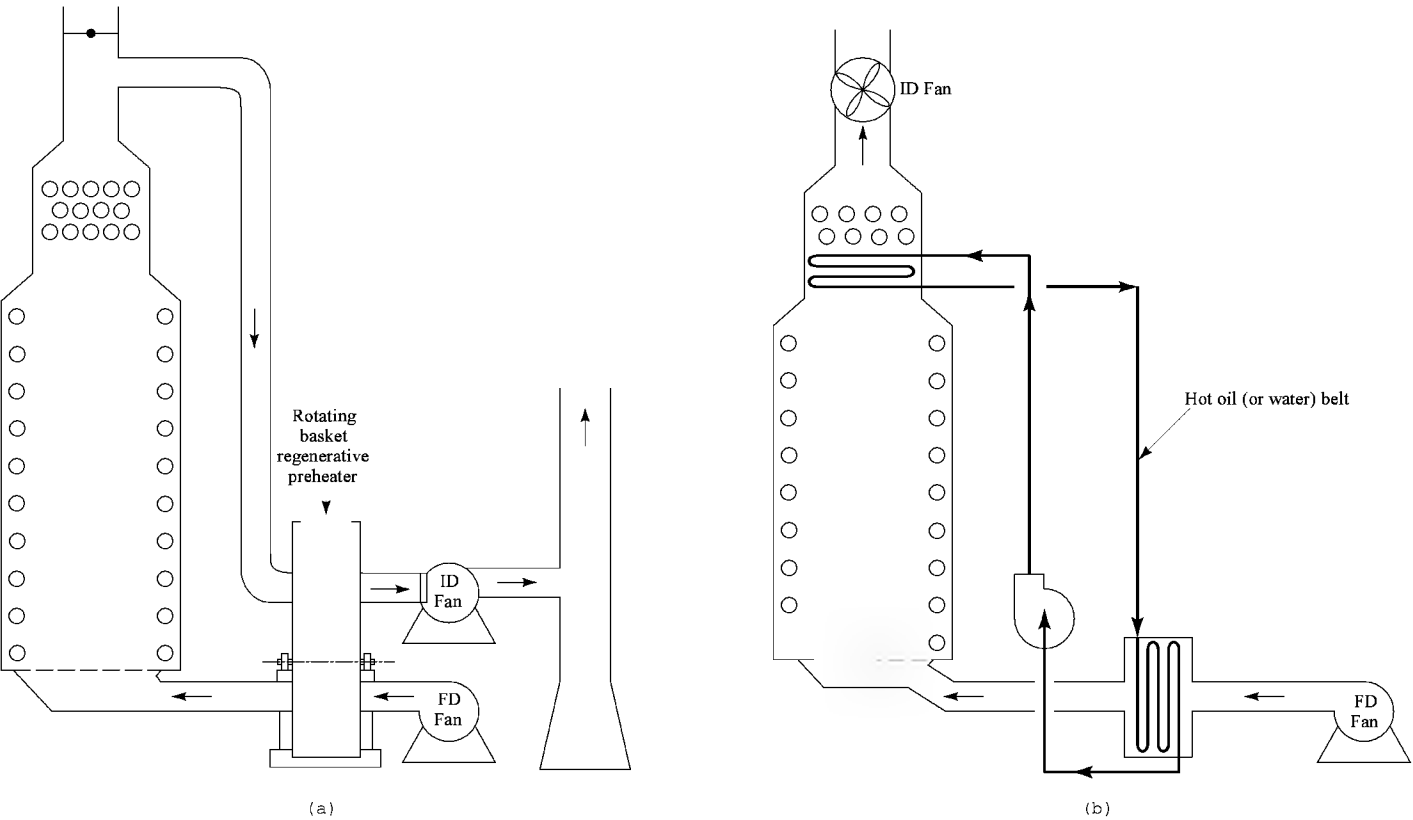


Fig. 7. Air preheaters where ID = induced draft and FD = forced draft fan. (a) Rotating metal basket or Lungstrom regenerative preheater; and (b) hot oil or water belt (liquid runaround) used to move convection section heat to air preheater in furnace retrofit.

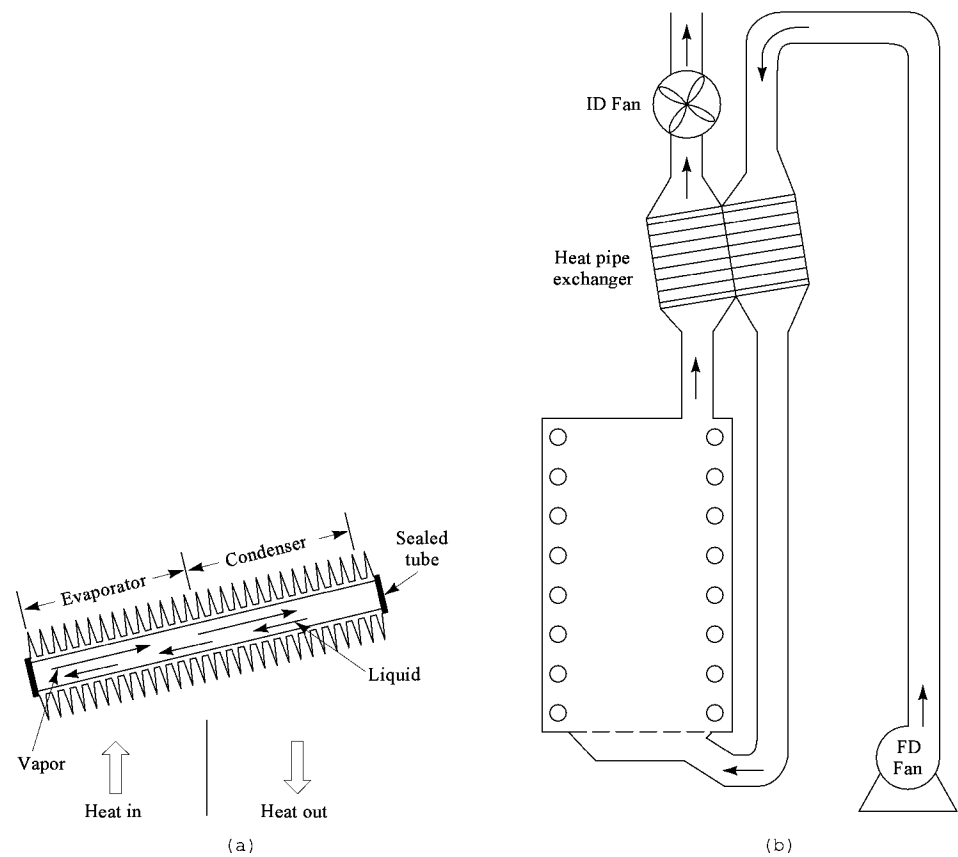


Fig. 8. (a) Heat pipe showing the use of finned tubing for both heating and cooling; and (b), heat pipe exchanger air heater system. ID = induced draft and FD = forced draft fan.

The heat pipe exchanger is a variant on the liquid runaround system where each tube (pipe) is sealed on both ends and filled with a vaporizing–condensing heat-transfer medium. At the hot end of the pipe, liquid is vaporized and moves to the cold end. At the cold end, vapor condenses and returns to the heat intake end. The flows are driven by gravity and capillary wicking. The heat pipe is particularly useful because it permits very compact, side by side ducting arrangements with countercurrent flow, as shown in Figure 8 (see HEAT EXCHANGE TECHNOLOGY, HEAT PIPES).

Boiler Economizers. Heat exchangers that use boiler flue gases to preheat the boiler feedwater are termed boiler economizers.

Heat Pumps. The use of heat pumps adds a compressor to boost the temperature level of rejected heat. It can be very effective in small plants having few opportunities for heat interchange. However, in large facilities a closer look usually shows an alternative for use of waste heat. The fuel steam focus of energy use has led to application of heat pumps in applications where a broader examination might suggest a simpler system of heat recovery.

Heat Recovery Equipment. Factors that limit heat recovery applications are corrosion, fouling, safety, and cost of heat-exchange surface. Most heat interchange utilizes shell and tube-type units because of the rugged construction, ease of mechanical cleaning, and ease of fabrication in a variety of materials. However, there is a rich assortment of other heat exchangers. Examples found in chemical plants in special applications include the following.

Plate heat exchangers are made by sandwiching thin sheets of metal that have a corrugated pattern pressed into them. The corrugations provide mechanical support where the sheets contact each other, permitting compact, low cost construction and generating high turbulence and high heat-transfer coefficients. These plates are normally separated by gaskets, and are particularly useful in food processing (qv), where cleaning is facilitated. However, the gaskets are the mechanical weak link, and limit the application in chemical plants to nontoxic, noninflammable fluids.

Brazed-fin aluminum cores are made of aluminum sheets having corrugated, cut layers sandwiched between flat sheets. The package is brazed together to form a very compact unit. The cut corrugated layers act as fins for both sides of the exchanger. The brazed aluminum construction needs to be protected from fire by special insulation or a coldbox filled with perlite. These units are used almost exclusively for very clean, cryogenic services such as air separation or hydrogen purification.

In spiral plate construction, two plates are welded together and rolled into a jelly-roll shape. The prime advantage is that there is a single flow passage. Any pluggage generates a high local pressure drop and tends to erode the deposit. Thus the unit is less subject to pluggage than in the other constructions having parallel flow paths.

Utility Systems

Steam. The steam system serves as the integrating energy system in most chemical process plants. Steam holds this unique position because it is an excellent heat-transfer medium over a wide range of temperatures. Water gives high heat-transfer coefficients whether in liquid phase, boiling, or in condensation. In addition, water is safe, nonpolluting, and if proper water treatment is maintained, noncorrosive to carbon steel.

Steam Balances. The steam balance is usually the most important plant-wide energy balance. It shows each service requirement, including the use of steam as a working fluid to develop power.

There are, however, some limitations. A steam balance depicts the steam flows at a given point in time or as an average over some period. The steam flow on a cold weekday morning in winter is quite different from that on a Sunday in summer, and neither matches the annual average steam balance. Startup flows are also usually far different and merit their own special balance. It is also wise to prepare a balance for the beginning and the end of the cycle between unit shutdowns. For example, the power required by a turbine driving a compressor rises as the compressor efficiency falls, and process heating requirements rise as interchangers foul. Use of a computer-based model permits easy adaptation to the calendar and onstream cycle.

There are two ways of presenting steam balance data, schematically or tabularly. For both presentation types, a balance is made at each pressure level. In a schematic balance, such as that shown in Figure 9, horizontal lines are drawn for each pressure. The steam-using equipment is shown between the lines, and individual flows are shown vertically. Table 3 contains the same data as shown in Figure 9. In both cases the steam balance has been simplified to show only mass flows. A separate balance should be developed that identifies energy flows, including heat losses and power extraction from the turbines.

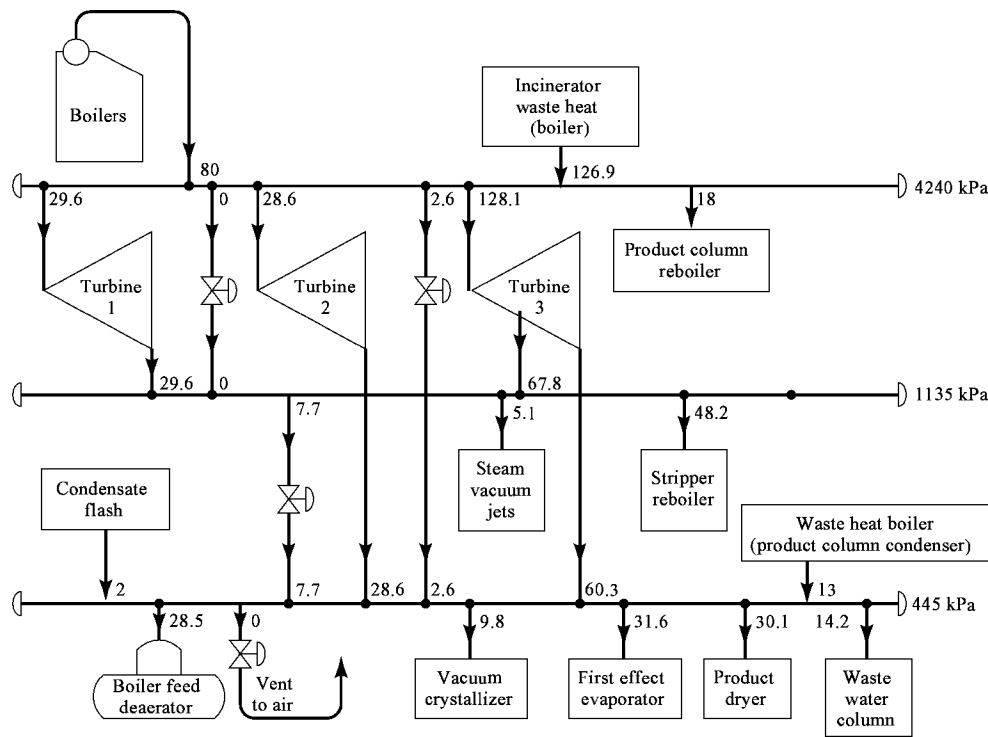


Fig. 9. Schematic steam balance where the numbers represent steam flows in metric tons per hour. See Table 3.

Table 3. Steam Balance, Flows in Metric Tons per Hour^a

Equipment	4240 kPa ^b		1335 kPa ^b		445 kPa ^b	
	Supply	Use	Supply	Use	Supply	Use
boilers	80.0					
incinerator waste heat	126.9					
turbine 1		29.6	29.6			
turbine 2		28.6			28.6	
turbine 3		128.1	67.8		60.3	
pressure reducing valve		0	0			
pressure reducing valve		2.6			2.6	
pressure reducing valve				7.7	7.7	
product column reboiler		18.0				
steam vacuum jets				5.1		
stripper reboiler				48.2		
finishing column reboiler				36.4		
condensate flash					2.0	
product column condenser						
waste-heat boiler					13.0	
vacuum crystallizer						9.8
1st effect evaporator						31.6
product dryer						30.1
wastewater column						14.2
boiler feed deaerator						28.5
Total	206.9	206.9	97.4	97.4	114.2	114.2

^a See Figure 9.

^b To convert kPa to psi, multiply by 0.145.

Steam Turbines. Historically, back-pressure steam turbines were used as drives throughout processes to increase reliability to cover electrical power failures. A typical turbine would be a single-stage 375 kW machine having throttle steam at 4240 kPa (600 psig) and exhaust at 1135 kPa (150 psig). The turbine would be controlled by a centrifugal fly-weight governor operating a single throttle valve. The efficiency would be about 40% when operated at rated conditions; ie, for the amount of steam passing through the turbine, it would develop 40% of the power that could be developed by an ideal turbine, expanding the steam isentropically. The efficiency was substantially lower when the machine was operated at part load, because a large portion of the pressure drop at part load was lost across the throttling valve, producing no work.

Because of increased emphasis on maximizing cogenerated power, newer plants are trying to utilize back-pressure turbines only in applications where efficiencies above 70% can be attained. This typically means limiting the applications to the large (>1000 kW) drives, and using small machines only where they are necessary for the safe shutdown of the unit. Multistage turbines are used even on the smaller loads.

Most large plants also have some condensing turbines to handle process and seasonal swings and provide some flexibility to the steam balance. For large (>15,000 kW) applications, condensing turbines can compete with purchased electricity. For small applications, power can be provided at much lower cost by motors. Condensing turbines generally have high reliability, and are also used where the costs of electrical power failure, in process downtime, are high. Public utility plants usually have condensing turbines at the bottom of a power cycle as shown by Figure 6d.

Condensate Return Systems. In a process plant, steam traps are used to drain and return condensate. Given proper application and continuous maintenance, these can operate with minimal steam leakage. Correct installation is also important (12).

For draining principal items of process equipment, level-controlled condensate chambers provide much better performance and reliability than steam traps. Usage is generally justified when condensate flow is greater than 4500 kg h.

Electrical. The plant electrical system is sometimes more important than the steam system. The electrical system consists of the utility company's entry substation, any in-plant generating equipment, primary distribution feeders, secondary substations and transformers, final distribution cables, and various items of switch-gear, protective relays, motor starters, motors, lighting control panels, and capacitors to adjust power factor.

Electric Motors. Except for electrolytic, eg, chlorine production (see ALKALI AND CHLORINE PRODUCTS; ELECTROCHEMICAL PROCESSING) or electric furnace processes, eg, phosphorus (see PHOSPHORUS AND THE PHOSPHIDES), typically 95% of the electricity used in a chemical plant is for electric motor drives. Induction electric motors in general use in chemical process plants range in efficiency from 90 to 95% depending primarily on size. The larger motors are more efficient. For any size, a range of efficiencies is available and high efficiency motors are somewhat more expensive than standard ones as shown in Figure 10. This price increment is normally justified in chemical process plants because of the high number of annual operating hours (13). For heating and ventilating operations, the lower operating hours sometimes make the lower efficiency units the cost-effective choice.

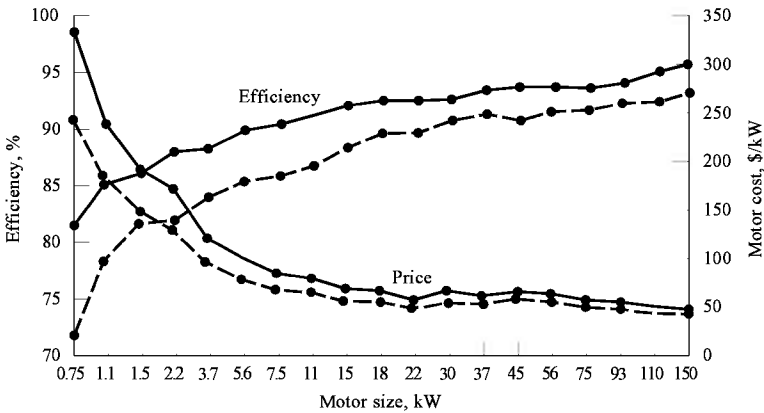


Fig. 10. Full-load efficiencies and wholesale prices (in 1988 U.S. dollars) of () standard and (—) energy efficient (EEM) three-phase motors (13).

Variable Frequency Drives. An important energy by-product of solid-state electronics is the relatively low cost variable speed drive. These electronic devices adjust the frequency of current to control motor speed such that a pump can be controlled directly to deliver the right flow without the need for a control valve and its inherent pressure drop. Figure 11 shows that at rated load the variable speed drive uses only about 70% as much power as a standard throttle control valve system, and at half load, it uses only about 25% as much power.

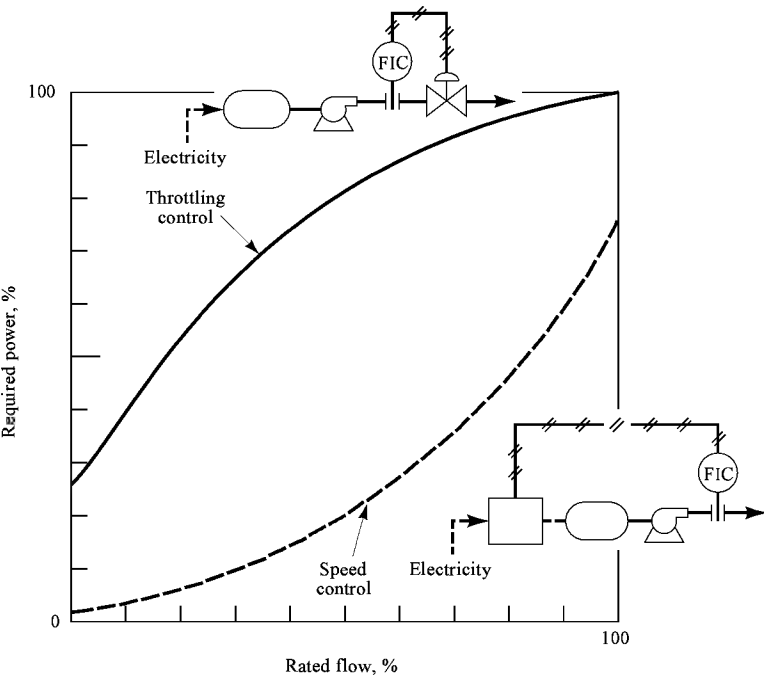


Fig. 11. Power saving for variable speed drives. Power input for variable speed adjusts with flow to naturally match the frictional losses. FIC = flow indicating controller.

In addition to energy conservation, the variable speed drives offer better control because of a faster response, ie, reduced dead band. They are also sometimes chosen for safety reasons because of elimination of the control station and accompanying valving. The capital saved by use of a smaller motor and elimination of the control valve partially compensates for the cost of the drive.

Lighting. High pressure sodium or metal halide lights offer significant savings over mercury vapor for process areas. Fluorescent lights have an even greater savings over incandescent lights for areas having low (mounting height <4 m) ceilings. Some other features that minimize electrical use in lighting are circuit arrangement so that unneeded lights can be turned off during non-use hours, and efficient reflectors and ballasts.

Other Energy Systems. Chemical plants usually require cooling water, compressed air, and fuel distribution systems. Sometimes also included are refrigeration, pressurized hot water, or specialized heat-transfer fluids such as Therminol liquid or condensing vapor. Each of these systems serves the process and reliability is the most important characteristic. Thus a project in any of them that achieves a 10% reduction in energy cost at the expense of a 1% loss of reliability loses money for the operation.

Cooling Water. The primary reliability concern is that water chemistry must be maintained in a low fouling, noncorroding regime. In addition, water flow velocity must be maintained above a certain threshold (ca 0.5 m/s in tubed side flow) to avoid fouling and corrosion.

The principal energy cost concern is avoidance of flows greatly in excess of need. Some cooling water systems are operated at such low temperature rises and high flows that pumping power cost equals the cost of the original heat input that it rejects. Care should be taken in design to assure that the system is balanced, and all heat exchangers utilize available pressure drop with the design flow. Every user should have a thermowell in the cooling water outlet to monitor temperature rise, and enough flow and temperature measuring elements should be provided to check overall heat balances (see TEMPERATURE MEASUREMENT).

Compressed Air. Enough flow elements should be provided to permit accurate assessment of use by different operating departments within a plant. Compressed air is often unmetered; thus there is little motivation to reduce use. A large fraction is often lost through leakage at fittings.

A key question for energy reduction is whether a lower discharge pressure can be used. Dropping pressure from 790 kPa (100 psig) to 650 kPa (80 psig) reduces the required power 12%, and reduces the driving force for air leakage losses by 20%. Controls such as inlet guide vanes on the air compressor can be provided to trim pressure to the required level. Another action that lowers energy use is lowering inlet and interstage temperatures. A drop from 30

to 20°C typically reduces power by 3^o %.

Refrigeration. In processes such as olefin separations, the economic importance of refrigeration exceeds that of the steam system. Refrigeration is an extremely valuable utility because of the work required to raise the energy to ambient temperature. Its value goes up directly as the temperature gap between ambient and use level goes up. For example, whereas refrigeration at -25°C is worth approximately as much as heat at 250°C, refrigeration at -75°C is worth twice as much.

A series of energy audit questions asking what can be done to reduce the work of raising this heat to ambient, or whether refrigeration is really necessary, should be addressed. Energy costs are cut if cooling water can be used directly or used for part of the year or for part of the load. An energy savings can also be realized if a higher refrigeration temperature can be used. If the refrigeration need is above 5°C, and there is waste heat available above 90°C, an absorption refrigeration system can be used to supply refrigeration.

The optimization of heat-transfer surfaces also plays a role. At the optimum, the lifetime cost of a surface is approximately equal in value to the lifetime cost of power used to overcome the temperature differential in the condenser and evaporator. Additionally, condensation on insulation is a sign of questionable insulation (see INSULATION, THERMAL). Frost is a certain signal that insulation can be improved.

Condensing Organic Vapor. The eutectic mixture of diphenyl and diphenyl oxide is an excellent vapor medium for precise temperature control at temperatures higher than those practical using steam. This mixture can achieve 315°C while holding pressure at 304 kPa (3 atm) absolute. In contrast, steam would require 10.6 MPa (105 atm) pressure.

These systems, commercially known as Therminol VP or Dowtherm A, differ from steam in some key areas which can result in operating problems unless handled properly in design (14). The low pressure–high temperature operation means that the $\Delta T / \Delta P$ ratio at saturation is quite high; for example, at 315°C the ratio is 25 times that of steam. This means that a pressure drop that would be nominal in a steam system (10 kPa (0.1 atm)), can not be tolerated if precise temperature control is needed.

Another difference is that molecular weight is much higher than that of the common noncondensables, and hence the noncondensables are harder to purge. In contrast, in a steam system almost all noncondensables are heavier than steam and tend to flush out with the condensate.

Capital and Equipment Areas

Virtually all chemical processing is energy driven, but in separations such as distillation, drying (qv), and evaporation (qv), this is particularly clear. All three of these processes are simple thermal operations that involve separation through vaporization, and only a minor change in the chemical energy of the products. The capital related costs of those operations are typically three to five times as large on a lifetime basis as the energy costs. In almost all cases there is a balance between capital and energy costs, and typically one is traded against the other to achieve the lowest overall cost. Much can be said about this energy–capital trade based on first principles and engineering correlations (15).

Insulation. A surprisingly important capital element of energy management is insulation. On large projects the capital cost of insulation is in the same range as that for heat exchangers or distillation towers and trays. At the optimum insulation thickness, the lifetime value of the insulation approximates the lifetime value of the heat loss; ie, insulation is as costly as the heat loss that it prevents. Uninsulated flanges, when they exist, are a particularly severe loss point, and when flanges need to be opened periodically, insulation via removable blankets is usually justified.

Insulation provides other functions in addition to energy conservation. A key role for insulation is safety. It protects personnel from burns and minimizes hot surfaces that could ignite inflammables. It also protects equipment, piping, and contents in event of fire. Thus materials such as mineral wool are sometimes used despite relatively poor thermal qualities.

Corrosion under insulation is also a concern, particularly in refrigeration systems. The specification of the insulation system needs to include painting, vapor barriers, and external metal jackets (16).

Compressors. Compression equipment accounts for a large fraction of power use as well as a large fraction of installed capital. Usually the energy bill for a compressor is large enough to warrant a very visible monitor of the driver, such as a control room electric meter. Testing programs to ensure operation of the compressor and its driver at peak energy efficiency are also justified. Temperature rise across a compressor is a simple and effective way to monitor efficiency.

Compressors are relatively fragile, precision machines, and require more care in specification and maintenance than any other part of the process. The largest process compressors are axial flow machines, for example the air compressors for gas turbines. Centrifugal machines can also handle large volumes, for example the cracked gas compressors in ethylene plants. Centrifugal machines are a bit less efficient, but are more rugged and tolerant of fouling service. Smaller volume compressors are reciprocating or rotary designs.

Energy consumption is but one of the selection criteria. For example, a reciprocating compressor usually delivers 5 to 25^o % higher efficiency than a centrifugal machine. In the size range where a single unit compressor can handle the flow, this usually pays for increased maintenance, but it rarely justifies the increased capital of parallel units in competition with a larger single train centrifugal.

A compressor is typically a specially designed device, and comes with far less surplus capacity than other process components. As a result compressors merit great care in specification of flow, inlet pressure, and discharge pressure. Similarly, the control system and equipment need to be carefully matched to provide turndown with maximum efficiency.

Because of the large volumetric flow inherent in gases, the cost of power for incremental pressure drop is high. To a first approximation, incremental power is proportional to the volumetric flow of suction or discharge, multiplied by incremental pressure drop; hence the high importance of pressure drop optimization for the associated piping and heat exchangers. Volumetric flow also varies with the absolute value of temperature; hence suction cooling is another area for optimization. The drive for lower temperatures also provides the incentive for adding compression stages, with intercooling between stages.

Pumps. Energy use for pumps (qv) can best be controlled by design for the proper flow and discharge pressure. Constant speed electric motor driven pumps, having a large margin of safety on flow, are particularly wasteful. One solution is use of the variable speed drive. Another solution in an operating unit is trimming of oversized impellers.

Vacuum Systems. The basic question in vacuum systems is what can be done to cut design inert loading. Historically, inert flows were whimsically overspecified, and the vacuum systems were run until mechanical deterioration brought the system capacity down to the actual inert loading. One factor driving toward greater use of vacuum pumps is the large reduction they achieve in effluent to be treated (see VACUUM TECHNOLOGY).

Boilers and Process Furnaces. Boilers and process fired heaters are the entry point for the energy released from burning fuel. Fuel combustion is irreversible, and fired heaters are typically the principal loss point for work potential (6). The high irreversibility results from taking the chemical energy of fuel and degrading it to heat. Air preheat cuts energy losses by cutting fuel firing and increasing the flame burst temperature.

A more obvious energy loss is the heat to the stack flue gases. The sensible heat losses can be minimized by reduced total air flow, ie, low excess air operation. Flue gas losses are also minimized by lowering the discharge temperature via increased heat recovery in economizers, air preheaters, etc. When fuels containing sulfur are burned, the final exit flue gas temperature is usually not permitted to go below about 100°C because of severe problems relating to sulfuric acid corrosion. Special economizers having Teflon-coated tubes permit lower temperatures but are not commonly used.

Inadequate mixing of air and fuel can result in unburned combustibles in the flue gases. This results in energy losses, environmental problems, and damage from afterburning in the convection section. Heat leakage through refractories (qv) constitutes still another type of energy loss from combustion equipment. Because of the high temperature, the heat leakage is higher than on most equipment and can represent as much as 3^o % of fired fuel. On newer designs this value is typically closer to 1^o %.

Distillation.
Reflux Rate. The optimum reflux rate for a distillation column depends on the value of energy, but is generally between 1.05 times and 1.25 times the reflux rate, which could be used with infinite trays. At this level, excess reflux is a secondary contributor to column inefficiency. However, when designing to this tolerance, correct vapor–liquid equilibrium data and adequate controls are essential.

Control. Energy savings for improved control are surprisingly high. A 2 to 20^o % savings from a series of control projects has been reported (17).

A reflux reduction of 15° is typical. Improved control achieves this by permitting a reduction in the margin of safety that the operators use to handle changes in feed conditions. The key element is the addition of feed-forward capability, which automatically handles changes in feed flow and composition. One of the reasons for increased use of features such as feed-forward control is the reduced cost of computers and online analyzers.

Reboilers and Condensers. The real work used in a distillation varies with the temperature difference between the heating medium and the cooling medium. Part of this differential is the difference in boiling points between the overhead stream and the bottom stream. However, a larger portion often results from the temperature difference used in the reboiler and condenser. The optimum differential is generally under 20°C, and if refrigeration is used, the optimum can be as small as 3°C. A signal that an excessive temperature difference may exist is a condenser or reboiler having a shell diameter less than one-third the diameter of the column.

There are a number of ways to provide the heating or cooling medium at temperatures closer to the optimum level. One is by use of double-effect distillation, which uses the overhead vapor from one column as the heat source for another column such that the second column's reboiler becomes the first column's condenser. This basically cuts the temperature differential in half, and shows up as an energy saving because external heat is supplied to only one of the units.

Column Pressure Drop. Another element that sets the temperature differential across the distillation is the pressure drop in the column and its auxiliaries. One of the more recent changes is the introduction of special structured packings (see DISTILLATION) which give extremely low (10° of an equivalent column with trays) pressure drop. This energy benefit can show up in an overhead temperature high enough to permit generation of by-product steam. It can also show up in a variety of other ways including lower bottoms temperature, yielding less fouling and product degradation to by-products, as in the styrene-ethylbenzene separation.

The Overall Process. Each distillation column should be examined in context with the rest of the process as well as by itself. For example, an opportunity for energy saving may be a reduction of process recycles via a purer overhead or bottoms stream. Lowering or raising column pressure to facilitate interchange with other parts of the process is another possible opportunity.

Drying. A typical dryer mass and energy balance (Fig. 9, Table 3) shows that the heat loss is 10° of the fuel input. Improving insulation is one of the simplest ways to reduce energy input. Another simple way to reduce energy input is improving the dewatering (qv) of the feed. There is a great difference in energy input for dewatering as compared to the subsequent drying step, as shown by Figure 12.

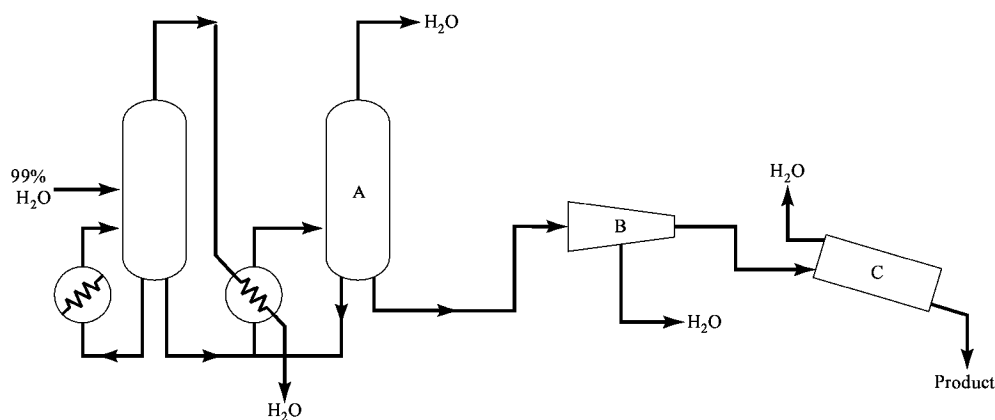


Fig. 12. The relative energy input for removal of 1 kg of water relative to heat of vaporization is 0.7 in A, the double-effect crystallizer; 0.015 in B, the dewatering centrifuge; and 5 in C, the rotary dryer.

Some of the other energy conservation approaches applicable to dryers are interchange between the stack vapor and the incoming dryer air; recovering sensible heat from the product; use of waste heat from another operation for air preheat; and using less, but hotter drying air. This last is limited to nonheat-sensitive materials.

Evaporation. In most chemical industry evaporation systems, the objective is product recovery, although occasionally the objective is concentration of an organic waste from an aqueous solution, to facilitate incineration. Similar equipment is used extensively for desalination of salt or brackish water (see also WATER, SUPPLY AND DESALINATION).

A single-effect evaporator produces slightly less than a kilogram of water vapor per kilogram of steam. By using the vapor produced by the first-effect as the heat source for a second-effect evaporator, steam use can be essentially halved. The performance can be improved almost in proportion to the number of effects employed. Six- and seven-effect evaporators are common in the wood pulp (qv) industry for concentration of black liquor. However, as the number of effects goes up, the temperature driving force is spread over the additional units and the capital cost increases almost in direct proportion to the number of effects. As a consequence, high alloy systems are often limited to single- or double-effect.

Much as reverse osmosis (qv) can compete with evaporation in desalination applications, osmosis should also be considered as an alternative for process evaporation. Reverse osmosis is particularly attractive where the inlet stream is greater than 99° water.

Energy Accounting and Improvement

Energy Accounting. Long-term costs at plant gate meters are costs that should be considered in project evaluations. The real benefit of a technical action can be quite different from the savings allocated by the accounting system. The benefit also depends on other changes that happen at the plant site. For example, the incremental value of steam may be near zero because of production from waste heat boilers and utility boilers operating at minimum load. Prudent planning credits a steam saving project based on the probable plant energy balance during the project's operation rather than on current allocated cost.

The incremental cost of electricity is usually dictated by contract. Incremental cost is normally less than the average cost because most utilities have descending rate scales for high volume industrial customers and also because there is a large fixed charged associated with past peak demand. An important factor in the control of operations is recognizing that the cost of electricity is typically one-third for peak demand and two-thirds for the actual incremental use. The demand charge is typically set by the peak usage in a month or 12 month period. The demand charge recognizes that a large component in the cost of electricity is the capital used by the public utility to generate and transmit power. Many chemical plants have load shedding systems to avoid setting a higher peak and increasing demand charges. Usually the shed load is for nonprocess applications such as battery charging of forklift trucks, but sometimes large users such as electric furnaces need to be curtailed as well.

Because it is so readily measurable, energy conservation offers a unique opportunity for continuous improvement. Energy conservation is something that individuals want to do, and given the opportunity to see ideas converted into measurable progress, personnel pursue conservation with enthusiasm, generating larger savings than deemed possible (18).

Measurements and Audits. The enabling element of continuous improvement is measurement. An old rule of thumb says that increased accuracy in measuring an energy use ultimately yields a reduction in use equal to 10° of the increased closure of the balance. A basic principle of economics is that any thing that is free is used in excess, ie, an unmetered electrical use is bigger than expected by at least 10°. Metering of the cost elements at each unit in a chemical plant provides effective accountability. Measurements should be linked via computer software to production as well as to weather to result in maximum feedback.

BIBLIOGRAPHY

"Energy Management" in *ECT* 3rd ed., Vol. 9, pp. 21–45, by A. F. Waterland, Waterland, Viar & Associates, Inc.

1. *Manufacturing Energy Consumption Survey: 1988*, U.S. Department of Energy, Washington, D.C., 1991.
2. *U.S. Chemical Industry Statistical Handbook*, Chemical Manufacturers Association, Washington, D.C., 1991.
3. W. M. Weiss, personal communication, Monsanto Enviro-Chem Systems Inc., St. Louis, Mo., 1991.
4. M. H. Ross and D. E. Steinmeyer, *Sci. Am.* **163**, 88 (1990).
5. *Manufacturing Energy Consumption Survey: 1985*, U.S. Department of Energy, Washington, D.C., 1988.
6. W. F. Kenney, *Energy Conservation in the Process Industries*, Academic Press, Inc., New York, 1984.
7. J. M. Kovacek, *Cogeneration Application Considerations*, General Electric Co., Schenectady, N.Y., GER-3430B, 1991.
8. W. F. Kenney, "Combustion Air Preheat on Steam Cracker Furnaces," *Proceedings, 1983 Industrial Energy Conservation Technology Conference*, Texas Industrial Commission, p. 595.
9. R. Smith and B. Linnhoff, *Chem. Eng. Res. Des.* **66**, 195 (1988).
10. P. S. Gupton and A. S. Krisher, *Chem. Eng. Prog.* **69**(1), 47 (1973).
11. V. Ganapathy, *Waste Heat Boiler Deskbook*, Prentice Hall, Inc., Englewood Cliffs, N.J., 1991.
12. F. S. Pychewicz, "Steam Traps—The Oft Forgotten Energy Conservation Treasure," *Proceedings, 1985 Industrial Energy Conservation Technology Conference*, Texas Industrial Commission, Houston, Tex., p. 392.
13. E. D. Larson and L. J. Nilsson, *ASHRAE Trans.* **97**(2), 363 (1991).
14. D. R. Frikken, K. S. Rosenberg, and D. E. Steinmeyer, *Chem. Eng.*, 86 (June 9, 1975).
15. D. E. Steinmeyer, *CHEMTECH* **188** (Mar. 1982).
16. L. G. Britton and H. G. Clem, *Chem. Eng. Prog.* **87**(11), 87 (1991).
17. T. L. Tolliver, *Chem. Eng.*, 99 (Nov. 24, 1986).
18. K. Nelson, *Chem. Proc.*, 77 (Jan. 1989).

Dan Steinmeyer
Monsanto Company

ENGINEERING, CHEMICAL DATA CORRELATION

It is a familiar scenario in the chemical sciences: a straightforward set of calculations becomes bogged down for lack of one or more pieces of data, particularly physical property data. Whereas the use of experimentally obtained data is usually preferred over values obtained from data prediction methods, such data are often unavailable or are of dubious quality. In such cases, correlations to obtain the required values must be employed. Because it is not sufficient to provide just a solution, but the best solution given the constraints in time, funding, and information, a large portion of scientific and engineering investigation is devoted to the assessment of the degree to which mathematical constructs faithfully mimic real systems. It is perhaps equally important to be able to determine the level of confidence to which a solution is ascribed.

Theoretically based correlations (or semitheoretical extensions of them), rooted in thermodynamics or other fundamentals are ordinarily preferred. However, rigorous theoretical understanding of real systems is far from complete, and purely empirical correlations typically have strict limits on applicability. Many correlations result from curve-fitting the desired parameter to an appropriate independent variable. Some fitting exercises are rooted in theory, eg, Antoine's equation for vapor pressure; others can be described as being semitheoretical. These distinctions usually do not refer to adherence to the observations of natural systems, but rather to the agreement in form to mathematical models of idealized systems. The advent of readily available computers has revolutionized the development and use of correlation techniques (see CHEMOMETRICS; COMPUTER TECHNOLOGY; DIMENSIONAL ANALYSIS).

It is often advisable to begin a project not by attempting to solve the real problem, but rather by constructing an idealized model which can be rigorously solved, then extrapolating based on assumptions and sensitivity analysis, to an estimation of the real solution. The field of classical thermodynamics, developed to relate physical properties to one another, is such an idealized model, although it cannot be used directly to predict physical properties. Empirical modifications to theoretically based models are often reliable because these tend to apply over a broad range of conditions and qualitatively exhibit the correct behavior near physical boundaries and limits. Other techniques such as molecular thermodynamics and kinetic theory might allow direct predictions of physical properties, but such schemes are usually complex and difficult to apply. Molecular descriptors such as shape, size, the chemical moieties present, etc, provide more information for basing correlations. For engineering and physical property data correlations it is important to: (1) determine the accuracy of the experimental data to which the correlation is to be applied; (2) understand the limitations of the correlation technique employed; (3) minimize mathematical complexity (additional fitting parameters may improve data regressions, but may also generate local minima maxima which do not exist beyond the mathematical model); (4) perform sufficient sensitivity analyses to be able to assess the confidence level of final results; and (5) present solutions within the context of confidence levels and assumptions.

An overview of some basic mathematical techniques for data correlation is to be found herein together with background on several types of physical property correlating techniques and a road map for the use of selected methods. Methods are presented for the correlation of observed experimental data to physical properties such as critical properties, normal boiling point, molar volume, vapor pressure, heats of vaporization and fusion, heat capacity, surface tension, viscosity, thermal conductivity, acentric factor, flammability limits, enthalpy of formation, Gibbs energy, entropy, activity coefficients, Henry's constant, octanol–water partition coefficients, diffusion coefficients, virial coefficients, chemical reactivity, and toxicological parameters. Correlation methods discussed include basic mathematical and numerical techniques, and approaches based on reference substances, empirical equations, nomographs, group contributions, linear solvation energy relationships, molecular connectivity indexes, and graph theory. Chemical data correlation foundations in classical, molecular, and statistical thermodynamics are introduced.

Thermodynamic Correlations for Property Estimation

In the broadest sense, thermodynamics is concerned with mathematical relationships that describe equilibrium conditions as well as transformations of energy from one form to another. Many chemical properties and parameters of engineering significance have origins in the mathematical expressions of the first and second laws and accompanying definitions. Particularly important are those fundamental equations which connect thermodynamic state functions to real-world, measurable properties such as pressure, volume, temperature, and heat capacity (1–3) (see also THERMODYNAMIC PROPERTIES).

The phase rule specifies the number of intensive properties of a system that must be set to establish all other intensive properties at fixed values (3), without providing information about how to calculate values for these properties. The field of applied engineering thermodynamics has grown out of the need to assign numerical values to thermodynamic properties within the constraints of the phase rule and fundamental laws. In the engineering disciplines there is a particular demand for physical properties, both for pure fluids and mixtures, and for phase equilibrium data (4,5).

Data compilations, the first recourse for an engineering calculation requiring physical property or parameter data, are often incomplete or do not contain data within the appropriate range of temperature or pressure (6–9). For this reason, correlation and estimation methods play an important role in applied thermodynamics.

The common theme in the evolution of methods for property and parameter prediction is the development of equations, either theoretical or empirical, containing quantities that can be calculated from theoretical considerations or experimental data. Mathematical expressions for correlating thermodynamic data may take several forms.

Limiting Laws. Simple laws that tend to describe a narrow range of behavior of real fluids and substances, and which contain few, if any, adjustable parameters are called limiting laws. Models of this type include the ideal gas law equation of state and the Lewis-Randall fugacity rule (10).

Equations of State. Equations of state having adjustable parameters are often used to model the pressure–volume–temperature (PVT) behavior of pure fluids and mixtures (1,2). Equations that are cubic in specific volume, such as a van der Waals equation having two adjustable parameters, are the mathematically simplest forms capable of representing the two real volume roots associated with phase equilibrium, or the three roots (vapor, liquid, solid) characteristic of the triple point.

Fundamental Property Relation. The fundamental property relation, which embodies the first and second laws of thermodynamics, can be expressed as a semiempirical equation containing physical parameters and one or more constants of integration. All of these may be adjusted to fit experimental data. The Clausius-Clapeyron equation is an example of this type of relation (1–3).

Semiempirical Relationships. Exact thermodynamic relationships can be approximated, and the unknown parameters then adjusted or estimated empirically. The virial equation of state, truncated after the second term, is an example of such a correlation (3).

Mathematical Consistency. Consistency requirements based on the property of exact differentials can be applied to smooth and extrapolate experimental data (2,3). An example is the use of the Gibbs-Duhem coexistence equation to estimate vapor mole fractions from total pressure versus liquid mole fraction data for a binary mixture.

Chemical Potential. Equilibrium calculations are based on the equality of individual chemical potentials (and fugacities) between phases in contact (10). In gas–solid adsorption, the equilibrium state can be defined in terms of an adsorption potential, which is an extension of the chemical potential concept to pore-filling (physisorption) onto microporous solids (11–16).

Generalized Correlations. Generalized correlations are often the only recourse when a property value cannot be determined from empirical correlations or by other means. Several powerful correlating techniques fall under this category, including the principle of corresponding states (3,17), reduced property models (1), and the Polanyi-type characteristic curve for microporous adsorbents (14).

This summary of correlating approaches is far from complete. The focus herein is primarily on pure components, although the phase equilibrium

models presented do account for mixture interactions (10). The literature should be consulted for electrolyte systems or transport parameters such as diffusion coefficients (1–3,18).

Theoretical and Empirical Correlating Forms.

Fundamental Property Relation. For homogeneous, single-phase systems the fundamental property relation (3), is a combination of the first and second laws of thermodynamics that may be written as

$$d(nU) = Td(nS) + Pd(nV) + \sum(\mu_i dn_i) \tag{1}$$

where U = internal energy, T = absolute temperature, S = entropy, P = pressure, μ_i = chemical potential, V = system volume, n_i = moles of i , and n = total moles in the system.

The definitions of enthalpy, H , Helmholtz free energy, A , and Gibbs free energy, G , also give equivalent forms of the fundamental relation (3) which apply to changes between equilibrium states in any homogeneous fluid system:

$$d(nH) = Td(nS) + (nV) dP + \sum(\mu_i dn_i) \tag{2}$$

$$d(nA) = -(nS) dT - Pd(nV) + \sum(\mu_i dn_i) \tag{3}$$

$$d(nG) = -(nS) dT + (nV) dP + \sum(\mu_i dn_i) \tag{4}$$

The chemical potential is a partial molar quantity defined by the last term in equation 4:

$$\mu_i = \left[\frac{\partial(nG)}{\partial n_i} \right]_{T,P,n_j} \tag{5}$$

where the subscripted term indicates that the moles of all chemical species other than n_i are held constant, and ∂ is a partial derivative.

Clausius-Clapeyron Equation. Derived from equation 1, the Clapeyron equation is a fundamental relationship between the latent heat accompanying a phase change and pressure–volume–temperature (PVT) data for the system (1):

$$\Delta H = T\Delta V \frac{dP^{sat}}{dT} \tag{6}$$

where ΔV = the volume change accompanying the phase change at temperature T ; ΔH = the latent heat of the phase change; and dP^{sat}/dT = the differential change of the saturation pressure with temperature.

The Clapeyron equation is most often used to represent the relationship between the temperature dependence of a pure liquid's vapor pressure curve and its latent heat of vaporization. In this case, dP^{sat}/dT is the slope of the vapor pressure–temperature curve, ΔV is the difference between the volume of the saturated vapor, V_g , and the saturated liquid, V_ρ , and ΔH^{vap} is the latent heat of vaporization. Commonly, V_ρ is small in comparison to V_g and the ideal gas law is assumed for the vapor phase.

If the latent heat of vaporization is then assumed to be constant over the temperature range of interest, equation 6 can be integrated to give the Clausius-Clapeyron expression:

$$\ln P^{sat} = \frac{\Delta H^{vap}}{RT} + I \tag{7}$$

Two empirical parameters are evident in equation 7, the heat of vaporization and the integration constant, I . Experimental data indicate that the linear relationship suggested by Clausius-Clapeyron may not be followed over a large temperature range (4); therefore additional adjustable parameters have been added to equation 7 to improve its correlating ability. The most prominent of these is the Antoine equation:

$$\log_{10} P^{sat} = A - \frac{B}{T + C} \tag{8}$$

where the constant A is related to the constant of integration, B is analogous to the latent heat term, and C is a correction factor added to the temperature T , usually in $^{\circ}\text{C}$.

Extensive tabulations of Antoine parameters are available for many chemicals of importance to engineers, chemists, and environmental scientists (9,19,20). Caution is in order when using tabulated Antoine constants because several forms of the correlating equation are found in the literature. In particular, there are variations in the sign before the second term, the units of temperature, and the use of natural or decimal logarithms of the vapor pressure.

The equation developed by the Design Institute for Physical Property Data (DIPPR) is another successful correlating tool for vapor pressure (4). It is an empirical extension of the Antoine equation and has two additional constants, D and E :

$$\ln P^{sat} = A + \frac{B}{T} + C \ln T + DT^E \tag{9}$$

Equations of State. An equation of state can be an exceptional tool for property prediction and phase equilibrium modeling. The term equation of state refers to the equilibrium relation among pressure, volume, temperature, and composition of a substance (2). This substance can be a pure chemical or a uniform mixture of chemicals in gaseous or liquid form.

Theoretical equation forms may be derived from either kinetic theory or statistical mechanics. However, empirical and semitheoretical equations of state have had the greatest success in representing data with high precision over a wide range of conditions (1). At present, theoretical equations are more limited in range of application than empirical equations. There are several excellent references available on the application and development of equations of state (2,3,18,21).

The Virial Expansion. Many equations of state have been proposed for gases, but the virial equation is the only one having a firm basis in theory (1,3). The pressure-explicit form of the virial expansion is

$$Z = \frac{PV}{RT} = 1 + B'P + C'P^2 + D'P^3 + \dots \tag{10}$$

where the ratio PV/RT is called the compressibility factor and is given the symbol Z . An equivalent expression for Z , explicit in volume, is

$$Z = 1 + \frac{B}{V} + \frac{C}{V^2} + \frac{D}{V^3} + \cdots$$

(11)

Equations 10 and 11 are known as virial expansions, and the coefficients $B', C, D', \dots$ and $B, C, D, \dots$ are called virial coefficients. For a given substance, these coefficients are functions of temperature only (1–3).

Statistical mechanics provides physical significance to the virial coefficients (18). For the expansion in $1/V$, the term B/V arises because of interactions between pairs of molecules (eq. 11), the term C/V^2 , because of three-molecule interactions, etc. Because two-body interactions are much more common than higher order interactions, truncated forms of the virial expansion are typically used. If no interactions existed, the virial coefficients would be zero and the virial expansion would reduce to the ideal gas law ($Z = 1$).

Liquid-Phase Models. Theoretical models of the liquid state are not as well established as those for gases; consequently, the development of general equations for the description of liquid-phase equilibrium behavior is not far advanced. Cubic equations of state give a qualitative description of liquid-phase equilibrium behavior, but do not generally yield good quantitative results (3). For engineering calculations, equations and estimation techniques developed specifically for liquids must normally be used.

The Tait equation for the pressure–volume behavior of liquids (1) correlates data accurately, and is expressed mathematically as

$$V = V_0 - D \ln \left(\frac{P + E}{P_0 + E} \right)$$

(12)

where the parameters D and E are empirical constants for a given temperature, and V_0 and P_0 are molar volume and pressure values at some reference state of the liquid. If sufficient data are available to allow determination of D and E , the Tait equation can represent the isothermal pressure–volume relation up to very high pressures (1).

Another empirical model for liquid pressure–volume behavior is the generalized equation for the molar volumes of saturated liquids given by the Rackett equation:

$$V^{\text{sat}} = V_c Z_c^{(1 - T_r)^{0.2857}}$$

(13)

which employs critical volume, V_c , the compressibility factor at the critical point, Z_c , and reduced temperature, T_r . The Rackett equation produces results accurate to about 2° (1).

Mixing Rules. Kay's method is a mixing rule for determining the pseudocritical temperature, T_{cm} , and pressure, P_{cm} , of a mixture of gases (1,3). In this method, pseudocritical values for mixtures of gases are calculated on the assumption that each component in the mixture contributes to the pseudocritical value in the same proportion as the number of moles of that component. The resulting values are referred to as linearly weighted mole-average pseudocritical properties of pressure, temperature, and volume. Kay's method is classified as a two-parameter rule because only P_c and T_c for each component are involved in the calculation of the compressibility factor, Z . Pseudoreduced properties of the mixture are calculated from the pseudocritical values from Kay's method, allowing the mixture to be treated as a pure fluid. Pseudoreduced properties then feed into an appropriate generalized correlation. If mixture data are scarce, Kay's rule is a good starting point. Other empirical mixing formulas may also be found in the literature (3).

Correlations for Enthalpy of Vaporization. Enthalpy or heat of vaporization, which is an important engineering parameter for liquids, can be predicted by a variety of methods which focus on either prediction of the heat of vaporization at the normal boiling point, or estimation of the heat of vaporization at any temperature from a known value at a reference temperature (5).

Enthalpies of vaporization, ΔH_n^{vap} , at the normal boiling point, T_b , can be estimated roughly with the simple rule of Pictet and Trouton (19) which state that the ratio of ΔH_n^{vap} to T_b should have the same value for all liquids:

$$\Delta H_n^{\text{vap}} / T_b = \text{constant} \approx 88 \text{ J / (K}\cdot\text{mol)} \text{ (21 cal / (K}\cdot\text{mol))}$$

(14)

It is equivalent to say that entropy of vaporization is a constant value for non-associating liquids. Associating liquids, eg, ammonia, water, methanol, and ethanol, do not obey the rule of Pictet and Trouton. Despite its simplicity, the Pictet-Trouton view of liquid vaporization (19) is an excellent example of the many rules of thumb that have been useful aids in engineering calculations for decades (5,7,8,9,21). However, proper application requires an understanding of the physical reasoning behind each rule.

An equation that yields values of ΔH_n^{vap} in units of kJ / mol to within about 2° is Chen's equation (1):

$$\Delta H_n^{\text{vap}} = \frac{T_b [0.0331 (T_b - T_c) + 0.0297 \log_{10} P_c - 0.0923]}{1.07 - T_b / T_c}$$

(15)

where T_b is the normal boiling point of the liquid in Kelvin, T_c is the critical temperature in Kelvin, and P_c is the critical pressure in kPa.

Another useful method for predicting the heat of vaporization at the normal boiling point is Riedel's equation (1):

$$\frac{\Delta H_n^{\text{vap}}}{T_n} = \frac{(9.079 \times 10^{-3} \ln P_c - 0.051)}{0.930 - T_{rn}}$$

(16)

where ΔH_n^{vap} is the molar latent heat of vaporization at the normal boiling point in kJ / mol, T_n is the reduced temperature at the normal boiling point, and P_c is the critical pressure in kPa (1).

Estimates of the latent heat of vaporization of a pure liquid at any temperature from a known value at a single temperature are possible by several methods. The known value may be experimental, or it may be estimated by equation 16. Of the methods proposed, the Watson correlation (1), which is both simple and reliable, has found the greatest acceptance:

$$\frac{\Delta H_2}{\Delta H_1} = \left(\frac{1 - T_{r2}}{1 - T_{r1}} \right)^{0.38}$$

(17)

where ΔH_2 is the heat of vaporization of a pure liquid at reduced temperature, T_{r2} , and ΔH_1 is the heat of vaporization of the same liquid at T_{r1} . The exponent is sometimes treated as an adjustable parameter in the Watson correlation (7). For example, various other values of the exponent are available (22).

Limiting Laws.

Ideal Gas Behavior. In 1787 it was demonstrated that the volume of a gas varies directly with temperature if the pressure remains constant. Other investigations determined complementary correlating relations from which the perfect or ideal gas law was drawn (1–3). Expressed mathematically, the ideal gas law is

$$pV = nRT$$

(18)

where p is the pressure exerted by the gas, V is the volume of the gas, n is the number of moles of gas, R is the universal gas constant, and T is the absolute temperature of the gas. Real gases closely obey this pressure–volume–temperature relationship at low pressures and high temperatures.

No real gas obeys equation 18 precisely over wide ranges of temperature and pressure. Under ordinary circumstances, lighter gases such as hydrogen, oxygen, and air obey the ideal gas law closely. Gases such as sulfur dioxide, carbon dioxide, and large hydrocarbons, however, deviate considerably from ideal gas behavior, particularly at high pressures and low temperatures.

The ideal gas law is an example of a correlating expression that comes directly from experimental observations, but has theoretical significance. Despite its simplicity, the ideal gas law is an excellent estimation tool. Often, it is the first approximation in systems involving real gases of all types. Unfortunately for liquids and solids no laws of such general utility are available.

Ideal Liquid Solutions. Two limiting laws of solution thermodynamics that are widely employed are Henry's law and Raoult's law, which represent vapor–liquid partitioning behavior in the concentration extremes. These laws are used frequently in equilibrium problems and apply to a variety of real systems (10).

The equilibrium partitioning of a chemical solute between a liquid and vapor phase is governed by Henry's law when the liquid mixture is very dilute in the solute. Henry's law generally is valid at concentrations below 0.01 mol /L of solution, although the upper limit can sometimes extend to 0.1 mol /L or higher (10). Over this concentration range, a direct proportionality, ie, Henry's constant, is observed between the partial pressure of the chemical in the gas phase and its mole fraction in the liquid phase. Henry's constant, when expressed in this way, has units of pressure (3).

In general, equality of component fugacities, ie, chemical potentials, in the vapor and liquid phases yields the following relation for vapor–liquid equilibrium:

$$y_i \phi_i P_T = x_i \gamma_i P_i^{\text{sat}} \phi_i^{\text{sat}} \pi_i$$

(19)

in which the Poynting correction factor, π_i , is defined as:

$$\pi_i = \exp\left(\frac{V_i^l (P_T - P_i^{\text{sat}})}{RT}\right)$$

(20)

and where y_i and x_i are the equilibrium mole fractions of i in the vapor and liquid phases, respectively, V_i^l is the liquid molar volume of pure i , ϕ_i is the vapor-phase fugacity coefficient, and P_T is the total pressure.

At system pressures up to several tens of MPa, the fugacity coefficients, ϕ_i and ϕ_i^{sat} , and the Poynting factor, π_i , are usually near unity. A simplified version of equation 19 can therefore be used for the majority of vapor–liquid equilibrium problems:

$$y_i P_T = x_i \gamma_i P_i^{\text{sat}}$$

(21)

In equation 21 the vapor phase is considered to be ideal, and all nonideality effects are attributed to the liquid-phase activity coefficient, γ_i . For an ideal solution ($\gamma_i = 1$), equation 21 becomes Raoult's law for the partial pressure, p_i , exerted by the liquid mixture:

$$p_i = y_i P_T = x_i P_i^{\text{sat}}$$

(22)

Equation 22 is a special application of the general Lewis-Randall ideal solution model (3,10) that is typically used for near-ambient pressures and concentrated nonpolar solutes.

Partial Molar Properties. The properties of individual components in a mixture or solution play an important role in solution thermodynamics. These properties, which represent molar derivatives of such extensive quantities as Gibbs free energy and entropy, are called partial molar properties. For example, in a liquid mixture of ethanol and water, the partial molar volume of ethanol and the partial molar volume of water have values that are, in general, quite different from the volumes of pure ethanol and pure water at the same temperature and pressure (21). If the mixture is an ideal solution, the partial molar volume of a component in solution is the same as the molar volume of the pure material at the same temperature and pressure.

Perhaps the most significant of the partial molar properties, because of its application to equilibrium thermodynamics, is the chemical potential, μ_i . This fundamental property, and related properties such as fugacity and activity, are essential to mathematical solutions of phase equilibrium problems. The natural logarithm of the liquid-phase activity coefficient, $\ln \gamma_i$, is also defined as a partial molar quantity. For liquid mixtures, the activity coefficient, γ_i , describes nonideal liquid-phase behavior.

Heat Capacities. The heat capacities of real gases are functions of temperature and pressure, and this functionality must be known to calculate other thermodynamic properties such as internal energy and enthalpy. The heat capacity in the ideal-gas state is different for each gas. Constant pressure heat capacities, C_p , for the ideal-gas state are independent of pressure and depend only on temperature. An accurate temperature correlation is often an empirical equation of the form:

$$C_p' = \alpha + \beta T + \gamma T^2$$

(23)

where α , β , and γ are constants characteristic of the gas considered and T is temperature (3). The prime superscript on C_p denotes the ideal-gas state.

For the ideal-gas state there is an exact relation between the constant pressure heat capacity and the constant volume heat capacity, C_v , via the ideal-gas constant, R .

$$C_p' = R + C_v'$$

(24)

From this equation, the temperature dependence of C_p is known, and vice versa (21). The ideal-gas state at a pressure of 101.3 kPa (1 atm) is often regarded as a standard state, for which the heat capacities are denoted by C_p^0 and C_v^0 . Real gases rarely depart significantly from ideality at near-ambient pressures (3); therefore, C_p and C_v usually represent good estimates of the heat capacities of real gases at low to moderate, eg, up to several hundred kPa, pressures. Otherwise thermodynamic excess functions are used to correct for deviations from ideal behavior when such situations occur (3).

When experimental data are not available, methods of estimation based on statistical mechanics are employed (7,19). Classical kinetic theory suggests a contribution to C_p^0 of $\frac{5}{2} R$ for each translational degree of freedom in the molecule, a contribution of $\frac{1}{2} R$ for each axis of rotation, and of R for each vibrational degree of freedom. A crude estimate of C_p^0 for small molecules can be obtained which neglects vibrational degrees of freedom:

$$C_p^0 = R(1 + g/2)$$

(25)

where g is the sum of the translational and rotational modes in the molecule. Equation 25 gives only rough estimates, so empirical correlations such as

equation 23 are recommended whenever possible (2). Ideal-gas heat capacities increase smoothly with increasing temperature toward an upper limit, which is reached when all translational, rotational, and vibrational modes of molecular motion are fully excited.

Mathematical Consistency Requirements. Theoretical equations provide a method by which a data set's internal consistency can be tested or missing data can be derived from known values of related properties. The ability of data to fit a proven model may also provide insight into whether that data behaves correctly and follows expected trends. For example, poor fit of vapor pressure versus temperature data to a generally accepted correlating equation could indicate systematic data error or bias. A simple semilogarithmic form, (eg, the Antoine equation, eq. 8), has been shown to apply to most organic liquids, so substantial deviation from this model might indicate a problem. Many other simple thermodynamics relations can provide useful data tests (1–5,18,21).

From the definition of a partial molar quantity and some thermodynamic substitutions involving exact differentials, it is possible to derive the simple, yet powerful, Duhem data testing relation (2,3,18). Stated in words, the Duhem equation is a mole-fraction-weighted summation of the partial derivatives of a set of partial molar quantities, with respect to the composition of one of the components (2,3). For example, in an *n*-component system, there are *n* partial molar quantities, *M_i*, representing any extensive molar property. At a specified temperature and pressure, only (*n* – 1) of these properties are independent. Many experiments, however, measure quantities for every chemical in a multicomponent system. It is this redundancy in reported data that makes thermodynamic consistency tests possible.

The well-known Gibbs-Duhem equation (2,3,18) is a special mathematical redundancy test which is expressed in terms of the chemical potential (3,18). The general Duhem test procedure can be applied to any set of partial molar quantities. It is also possible to perform an overall consistency test over a composition range with the integrated form of the Duhem equation (2).

In some cases, reported data do not satisfy a consistency check, but these may be the only available data. In that case, it may be possible to smooth the data in order to obtain a set of partial molar quantities that is thermodynamically consistent. The procedure is simply to reconstruct the total molar property by a weighted mole fraction average of the *n* measured partial molar values and then recalculate normalized partial molar quantities. The new set should always be consistent.

For interrelated properties, a method of assessing data quality is to see if data for those properties are consistent with theoretical equations. One example of such a test could be how well the slope of a semilogarithmic vapor pressure plot agrees with that chemical's measured heat of vaporization. Recall that the integrated Clausius-Clapeyron form (eq. 7) suggests that a semilogarithmic plot of vapor pressure versus inverse temperature should be a straight line having a slope equal to the heat of vaporization.

Many additional consistency tests can be derived from phase equilibrium constraints. From thermodynamics, the activity coefficient is known to be the fundamental basis of many properties and parameters of engineering interest. Therefore, data for such quantities as Henry's constant, octanol–water partition coefficient, aqueous solubility, and solubility of water in chemicals are related to solution activity coefficients and other properties through fundamental equilibrium relationships (10,23,24). Accurate, consistent data should be expected to satisfy these and other thermodynamic requirements. Furthermore, equilibrium models may permit a missing property value to be calculated from those values that are known (2).

Phase Equilibria and Chemical Potential.

Models of Phase Nonideality. The correlation and prediction of phase equilibria at low (near-vacuum) to moderate (several hundred kPa) pressures is typically based on a form of equation 19. The key is the degree to which the various parameters for phase nonideality are known or can be estimated.

Fugacity Coefficients. An exact equation that is widely used for the calculation of fugacity coefficients and fugacities from experimental pressure–volume–temperature (PVT) data is

$$\ln \hat{\phi}_i = \int_0^P (Z_i - 1) \frac{dP}{P} (\text{constant } T, x)$$

26

where ϕ_i = fugacity coefficient of *i* in a vapor mixture, and

$$Z_i = PV_i/RT$$

27

which leads to another form of equation 26:

$$\ln \hat{\phi}_i = \frac{1}{RT} \int_0^P \left(\frac{RT}{P} - V_i \right) dP (\text{constant } T, x)$$

28

Here a suitable equation of state is required to provide a mathematical expression for the mixture molar volume, $\overline{V}_i$. For some equations of state, it is better to use a form of equation 28 in which the integral is volume explicit (3). Note also that for an ideal gas $Z_i = \overline{Z}_i = 1$, and $\phi_i = \hat{\phi} = 1$.

Activity Coefficients. Activity coefficients in liquid mixtures are directly related to the molar excess Gibbs energy of mixing, ΔG^E , which is defined as the difference in the molar Gibbs energy of mixing between the real and ideal mixtures. It is typically an assumed function. Various functional forms of ΔG^E give rise to many of the different activity coefficient models found in the literature (1–3,18). Typically, the liquid-phase activity coefficient is a function of temperature and composition; explicit pressure dependence is rarely included.

There are many simple two-parameter equations for liquid mixture constituents, including the Wilson (25), Margules (2,3,18), van Laar (3,26), nonrandom two-liquid (NRTL) (27), and universal quasichemical (UNIQUAC) (28) equations. In the case of the NRTL model, one of the three adjustable parameters has been found to be relatively constant within some homologous series, so NRTL is essentially a two-parameter equation. The third parameter is usually treated as a constant which is set according to the type of chemical system (27). A third parameter for Wilson's equation has also been suggested for use with partially miscible systems (29,30,31). These equations all require experimental data to fit the adjustable constants. Simple equations of this type have the additional attraction of being useful for hand calculations.

Other algorithms are much more complex and require computer software, such as group contribution models for estimating liquid-phase activity coefficients. These models include analytical solution of groups (ASOG) (32), modified separation of cohesive energy density (MOSCED) (33,34), and UNIQUAC functional-group activity coefficient (UNIFAC) (28,35–40). Each model is applicable for some, but not all, chemical systems and conditions, and the choice of the functional form for the Gibbs excess energy leads to specialization and limitation.

The solvophobic model of liquid-phase nonideality takes into account solute–solvent interactions on the molecular level. In this view, all dissolved molecules expose microsurface area to the surrounding solvent and are acted on by the so-called solvophobic forces (41). These forces, which involve both enthalpy and entropy effects, are described generally by a branch of solution thermodynamics known as solvophobic theory. This general solution interaction approach takes into account the effect of the solvent on partitioning by considering two hypothetical steps. First, cavities in the solvent must be created to contain the partitioned species. Second, the partitioned species is placed in the cavities, where interactions can occur with the surrounding solvent. The idea of solvophobic forces has been used to estimate such diverse physical properties as absorbability, Henry's constant, and aqueous solubility (41–44). A principal drawback is calculational complexity and difficulty of finding values for the model input parameters.

The solvophobic model has been used to deduce a functional form for a Henry's constant correlation based on molecular connectivity index and polarizability (42). Accurate predictions are reported over a span of seven log units in Henry's constant. A reliable solvophobic model of aqueous solubility has also been reported (45,46).

Regular Solution Theory. The key assumption in regular-solution theory is that the excess entropy, S^E , is zero when mixing occurs at constant volume (3,18). This idea of a regular solution (26) leads to the equations:

$$G^E = \sum_k \left[x_k V_k (\delta_k - \bar{\delta})^2 \right] \tag{29}$$

and

$$\ln \gamma_i = - \frac{V_i}{RT} (\delta_i - \bar{\delta})^2 \tag{30}$$

where

$$\bar{\delta} = \frac{\sum_k (x_k V_k \delta_k)}{\sum_k (x_k V_k)} \tag{31}$$

In these equations, V_k is the molar volume and δ_k is the solubility parameter of the pure liquid k . The solubility parameter is the square root of an energy density, defined as

$$\delta_i = \left(\frac{\Delta U_i^{\text{vap}}}{V_i} \right)^{1/2} \tag{32}$$

where ΔU_i^{vap} is the internal energy change of vaporization of pure liquid i . Values are determined from heats of vaporization.

The solubility parameter, δ_i , is a function of temperature, but the difference $(\delta_i - \bar{\delta})$ is only weakly dependent on temperature. By convention, both δ_i and V_i are evaluated at 25°C and are treated as constants independent of both T and P . The activity coefficients given by equation 30 are therefore functions of liquid composition and temperature, but not of pressure.

A key feature of this model is that no data for mixtures are required to apply the regular-solution equations because the solubility parameters are evaluated from pure-component data. Results based on these equations should be treated as only qualitative. However, mixtures of nonpolar or slightly polar, nonassociating chemicals, can sometimes be modeled adequately (1,3,18). Applications of this model have been limited to hydrocarbons (qv) and a few gases associated with petroleum (qv) and natural gas (see GAS, NATURAL) processing, such as N₂, H₂, CO₂, and H₂S. Values for δ_i and V_i can be found in many references (1–3,7).

Thermodynamics of Vapor–Liquid Equilibrium. Assuming ideal vapor and choosing the Lewis-Randall standard state, a chemical distributed between a vapor and liquid in equilibrium often behaves according to:

$$y_i P_T = p_i = x_i \gamma_i P_i^{\text{sat}}, \tag{33}$$

where y_i is the mole fraction of i in the vapor phase; x_i is the mole fraction of i in the liquid phase; P_T is the total system pressure; p_i is the partial pressure of i in the vapor mixture; γ_i is the liquid-phase activity coefficient of i (Lewis-Randall basis); and P_i^{sat} is the saturation pressure of pure liquid i at the system temperature. Note that equation 33, which is based on equal fugacities, ie, equal chemical potentials, of a partitioned component in the vapor and liquid, neglects the Poynting and saturation fugacity coefficient corrections presented earlier (eq. 19). This equilibrium model, despite its simplicity, is suitable for many engineering calculations. Setting $\gamma_i = 1$ reduces equation 33 to Raoult's law (1–3,18).

Henry's Law Constant. Henry's law for dilute concentrations of contaminants in water is often appropriate for modeling vapor–liquid equilibrium (VLE) behavior (47). At very low concentrations, a chemical's Henry's constant is equal to the product of its activity coefficient and vapor pressure (3,10,48). Activity coefficient models can provide estimated values of infinite dilution activity coefficients for calculating Henry's constants as a function of temperature (35–39,49).

The short-cut technique frequently used to estimate the Henry's constant of a volatile substance in water is to calculate the ratio of the pure compound's vapor pressure to its aqueous saturation limit (23):

$$H_i = \frac{P_i^{\text{sat}}}{x_s} \tag{34}$$

where x_s is the ultimate solubility of the substance in water in mole fraction, P_i^{sat} is the vapor pressure of the pure liquid substance in kPa, and H_i is the Henry's constant in kPa. This assumes the miscibility of water in the volatile substance is negligible. A more rigorous solution, however, yields

$$H_i = \frac{(1 - x_w) P_i^{\text{sat}}}{x_s} \tag{35}$$

where x_w is the saturation limit of water in the substance in mole fraction.

As indicated by equation 35, the vapor pressure–solubility ratio (eq. 34), only provides accurate results when (1) reliable vapor pressure and aqueous solubility data are employed, (2) the quantity x_w is small enough to be neglected, and (3) Henry's law applies for the dissolved substance in water up to the saturation limit. If water miscibility in the organic is expected to be significant, equation 35 is much preferred, assuming miscibility data are available (42). This correction can amount to several percent even when the aqueous solubility is low. For example, water solubility in *n*-octanol is about 5% whereas the *n*-octanol solubility in water is in the parts per thousand range.

From fundamental thermodynamic relations, the temperature and pressure dependence of Henry's constant can be shown (18,50,51) to be:

$$d(\ln H_i) = \left(\frac{h_i - h_i^\infty}{RT^2} \right) dT + \left(\frac{V_i^\infty}{RT} \right) dP \tag{36}$$

where h_i is the molar enthalpy of i in the ideal gas state, h_i^∞ is the partial molar enthalpy of i at infinite dilution, and V_i^∞ is the partial molar volume of i at infinite dilution.

For experiments conducted at constant pressure, the second term in equation 36 disappears. The expression for the temperature dependence is then obtained by performing an indefinite integration on the remainder of the equation after assuming that the enthalpy change of volatilization, $(h_i - h_i^\infty)$, is constant with respect to temperature. The resulting equation is

$$\ln H_i = \frac{(h_i - h_i^\infty)}{RT} + I \tag{37}$$

where I is an integration constant.

By treating the quantities $(h_i - h_i^\infty)$, R and I as fitting parameters, H_i data obtained at various temperatures can be correlated to equation 37 using linear regression. The final equation for H_i has the form:

$$H_i = \exp \left[B - \left(\frac{A}{T} \right) \right]$$

(38)

where A and B are the "best-fit" constants for the experimental data. In environmental systems, where there are temperature gradients, the temperature in equation 38 can be taken to be the air–water interface temperature (48).

Thermodynamics of Liquid–Liquid Equilibrium. Phase splitting of a liquid mixture into two liquid phases (I and II) occurs when a single liquid phase is thermodynamically unstable. The equilibrium condition of equal fugacities (and chemical potentials) for each component in the two phases allows the fugacities f_i and f'_i in phases I and II to be equated and expressed as:

$$x'_i \gamma'_i = x''_i \gamma''_i$$

(39)

The same reference (standard) state, $f^\ominus$, is chosen for the two phases, so that it cancels on both sides of equation 39. The products $x'_i \gamma'_i$ and $x''_i \gamma''_i$ are referred to as activities. Because equation 39 holds for each component of a liquid–liquid system, it is possible to predict liquid–liquid phase splitting when the activity coefficients of the individual components in a multicomponent system are known. These values can come from vapor–liquid equilibrium experiments or from prediction methods developed for phase-equilibrium problems (4,5,10). Some binary systems can be modeled satisfactorily in this manner, but only rough estimations appear to be possible for multicomponent systems because activity coefficient models are not yet sufficiently developed in this area.

Aqueous Solubility. Solubility of a chemical in water can be calculated rigorously from equilibrium thermodynamic equations. Because activity coefficient data are often not available from the literature or direct experiments, models such as UNIFAC can be used for structure–activity estimations (24). Phase-equilibrium relationships can then be applied to predict miscibility. Simplified calculations are possible for low miscibility; however, when there is a high degree of miscibility, the phase-equilibrium relationships must be solved rigorously.

For low miscibility, the solubility of a substance in water is often estimated as the inverse of the chemical's infinite dilution activity coefficient:

$$x_s \simeq \frac{1}{\gamma_i^\infty}$$

(40)

where x_s = the solubility limit of the chemical in water in mole fraction, and γ_i^∞ = the infinite dilution activity coefficient, which is either known from experiments or estimated. Equation 40 is valid only when Henry's law applies to the dissolved chemical up to its aqueous solubility limit and the solubility of water in the chemical can be neglected.

If miscibility is significant in a binary chemical–solvent system, the calculations become more complex because the coupled nonlinear phase equilibrium expressions must be solved (10):

$$x_s \cdot \gamma_s^\text{II} = (1 - x_w) \cdot \gamma_s^\text{I}$$

(41)

$$x_w \cdot \gamma_w^\text{I} = (1 - x_s) \cdot \gamma_w^\text{II}$$

(42)

and

$$\gamma_i = f(x_s, x_w, T)$$

(43)

In the case of water as the solvent, I designates the chemical-rich phase, II corresponds to the water-rich phase, x_i is the aqueous solubility of the chemical, and x_w is the solubility of the water in the chemical. The activity coefficients are nonlinear functions of concentration, requiring a numerical solution to find the values of x_i and x_w that satisfy equations 41 and 42. In principle, this method can be extended to multiple components, and general-purpose "flash" procedures are available in the form of programmable algorithms (3,10). High quality organic-water mutual miscibility data are scarce (52–60), underscoring the need for predictive methods.

Octanol–Water Partition Coefficient. In environmental calculations, the octanol–water partition coefficient, K_{ow} , is related to a chemical's lipophilicity and correlates well with many properties. K_{ow} is defined as the ratio of the concentrations of a third component distributed between octanol-rich and water-rich phases in equilibrium. As the mole fraction of the distributed chemical in the two phases approaches zero, the K_{ow} concentration ratio approaches a constant value (23).

Based on liquid–liquid equilibrium principles, a general model of octanol–water partitioning is possible if accurate activity coefficients can be determined. First, phase equilibrium relationships based on activity coefficients permit liquid–liquid equilibrium calculations for the binary octanol–water system. Because the two components are almost immiscible in each other, two phases form: an octanol-rich phase containing dissolved water, and a water-rich phase containing dissolved octanol.

Once the composition of each equilibrium phase is known, infinite dilution activity coefficients for a third component in each phase can then be calculated. The octanol–water partition coefficient is directly proportional to the ratio of the infinite dilution activity coefficients for a third component distributed between the water-rich and octanol-rich phases (5,24). The primary drawback to the activity coefficient approach to K_{ow} estimation is the difficulty of the calculations involved, particularly when the activity coefficient model is complex.

Generalized and Reduced Equations

Reduced conditions are corrected, or normalized, conditions of temperature T , pressure p , and specific volume $V^\wedge$, and are expressed mathematically as

$$T_r = \frac{T}{T_c}$$

(44)

$$p_r = \frac{p}{p_c}$$

(45)

$$\hat{V}_r = \frac{\hat{V}}{\hat{V}_c}$$

(46)

where the subscript c denotes the critical state, and r the reduced state.

The idea of using reduced variables to correlate the pressure–volume–temperature properties of gases, was suggested by van der Waals in 1873. The

underlying principle is that all substances behave alike in the reduced, ie, corrected, states. That is, any substance would have the same reduced volume at the same reduced temperature and reduced pressure. Certainly, the simplest form of such an equation would be a variation of the ideal gas law:

$$p_r \widehat{V}_r = \gamma RT_r$$

(47)

where γ is some universal constant and $\widehat{V}_r$ is the reduced specific molar volume. Although equation 47 works for many gases at high temperature and pressure, it fails to predict gas properties adequately in the region where the ideal gas law applies, ie, at low pressure somewhat below $p_r = 1.0$.

Principle of Corresponding States.

Generalized Correlations. A simple and reliable method for the prediction of vapor–liquid behavior has been sought for many years to avoid experimentally measuring the thermodynamic and physical properties of every substance involved in a process. Whereas the complexity of fluids makes universal behavior prediction an elusive task, methods based on the theory of corresponding states have proven extremely useful and accurate while still retaining computational simplicity. Methods derived from corresponding states theory are commonly used in process and equipment design.

A generalized correlation is a functional relationship between dimensionless or reduced variables. Most generalized correlations may be expressed by the following function:

$$z = \psi(T_r, P_r)$$

(48)

which is the keystone to the theory of corresponding states. This form readily lends itself to the use of a generalized chart, where the compressibility factor (or any other property) is plotted as a function of the reduced pressure for each reduced isotherm. A great advantage of the generalized chart is its ease of use and fair accuracy (within 10%) obtained from limited information, ie, usually the critical pressure and critical temperature. Drawbacks to the generalized chart include the inability to predict properties with high precision (<1% error) and the inconvenience of the form to computer application.

Generalized charts are applicable to a wide range of industrially important chemicals. Properties for which charts are available include all thermodynamic properties, eg, enthalpy, entropy, Gibbs energy, and PVT data, compressibility factors, liquid densities, fugacity coefficients, surface tensions, diffusivities, transport properties, and rate constants for chemical reactions. Charts and tables of compressibility factors vs reduced pressure and reduced temperature have been produced. Data is available in both tabular and graphical form (61–72).

An alternative to the use of generalized charts is an analytical equation of state. Equations of state which are expressed as a function of reduced properties and nondimensional variables are said to be generalized. The term generalization is in reference to the wide applicability to the estimation of fluid properties for many substances.

Numerous other methods have been used to predict properties of gases and liquids. These include group contribution, reference substance, approaches, and many others. However, corresponding states theory has been one of the most thoroughly investigated methods and has become an important basis for the development of correlation and property estimation techniques. The methods derived from the corresponding states theory for liquid and gas property estimation have proved invaluable for work such as process and equipment design.

Reduced Properties. One of the first attempts at achieving an accurate analytical model to describe fluid behavior was the van der Waals equation, in which corrections to the ideal gas law take the form of constants a and b to account for molecular interactions and the finite volume of gas molecules, respectively.

$$\left(P + \frac{a}{\widehat{V}^2}\right)(\widehat{V} - b) = RT$$

(49)

This equation bridged the gap between gases and liquids by showing each could be described by a single cubic equation. Although the van der Waals equation's accuracy fails at the critical point and in the liquid region, it has had a great effect on the understanding of fluids. It is from this simple equation that the theory of corresponding states was proposed.

All fluids, when compared at the same reduced temperature and reduced pressure have approximately the same compressibility factor and deviate from ideal gas behavior to the same extent, giving

$$\left(P_r + \frac{3}{\widehat{V}_r^2}\right)(3\widehat{V}_r - 1) = 8T_r$$

(50)

which is theoretically equivalent for all substances.

In 1893, it was shown that corresponding states are not unique to van der Waals' equation of state (73). Rather, for any equation of state having not more than three constants, corresponding states are only a mathematical consequence.

The concept of corresponding states was based on kinetic molecular theory, which describes molecules as discrete, rapidly moving particles that together constitute a fluid or solid. Therefore, the theory of corresponding states was a macroscopic concept based on empirical observations. In 1939, the theory of corresponding states was derived from an inverse sixth power molecular potential model (74). Four basic assumptions were made: (1) classical statistical mechanics apply, (2) the molecules must be spherical either by actual shape or by virtue of rapid and free rotation, (3) the intramolecular vibrations are considered identical for molecules in either the gas or liquid phases, and (4) the potential energy of a collection of molecules is a function of only the various intermolecular distances.

Using these assumptions and a classical partitioning function, Q , integrated over the coordinates and momenta of all molecules, a universal function was defined:

$$Q = F\left(\frac{\widehat{V}}{\widehat{V}_c}, \frac{T}{T_c}\right)$$

(51)

Whereas this two-parameter equation states the same conclusion as the van der Waals equation, this derivation extends the theory beyond just PVT behavior. Because the partition function, Q , can also be used to derive all the thermodynamic functions, the functional form, F , can be changed to describe this data as well. Corresponding states equations are typically written with respect to temperature and pressure because of the ambiguities of measuring volume at the critical point.

The theory of corresponding states only works for certain classes of fluids, thus fluids have been placed into generalized groups based on behavior (75,76). These groupings include the following.

Simple Fluids. Spherical compounds having little molecular interaction, eg, argon, krypton, xenon, and methane, are known as simple fluids and obey the theory of corresponding states.

Normal Fluids. Asymmetrical compounds having little molecular interaction, eg, carbon monoxide, *n*-butane, and *n*-hexane, deviate slightly from the theory of corresponding states and are considered to be normal fluids.

Slightly Polar Fluids. Compounds having weak polarity such as benzene and toluene cannot be accurately predicted by corresponding states and are known as slightly polar fluids.

Hydrogen-Bonded, Associating, and Polar Fluids. Compounds having strong hydrogen-bonding and polar behavior, eg, water and alcohols, as well as compounds having strong ionic characteristics, such as weak acids and bases, are not applicable to simple corresponding states equations. Only complex

equations can begin to describe the behavior of these compounds.

Quantum Fluids. Light compounds which exhibit behavior resulting from quantum effects, eg, hydrogen, helium, and neon, are called quantum fluids.

Reduced Equations of State. A simple modification to the cubic van der Waals equation, developed in 1946 (72), uses a term called the ideal or pseudocritical volume, $V_c^{ideal} = RT_c/P_c$, to avoid the uncertainty in the measurement of volume at the critical point.

$$P_r = \frac{T_r}{0.125 - \phi} - \frac{0.422}{\phi^2}$$

(52)

where

$$\phi = \frac{\hat{V}}{V_c^{ideal}}$$

This expression, which has an average error of 2% or less, is valid up to the critical density. It also allows a corrective factor to be added to the critical pressure and temperature so that hydrogen and helium behavior can be accurately predicted (3,21).

A generalized cubic equation is the Redlich-Kwong equation (77):

$$Z = \frac{1}{1 - h} - \frac{4.9340}{T_r^{1.5}} \left(\frac{h}{1 + h} \right)$$

(53)

where

$$h = \frac{0.08664 P_r}{Z T_r}$$

This equation is useful for gases above the critical point. Only reduced pressure, P_r , and reduced temperature, T_r , are needed. In the form represented by equation 53, iteration quickly gives accurate values for the compressibility factor, Z. However, this two-parameter equation only gives accurate values for simple and nonpolar fluids. Unless the Redlich-Kwong equation (eq. 53) is explicitly solved for pressure in nonreduced variables, it does not give accurate liquid volumes.

An example of a more complex equation that exhibits better accuracy is the Benedict-Webb-Rubin equation:

$$P = \frac{RT}{\hat{V}} + \frac{B_0 RT}{\hat{V}^2} - \frac{A_0}{\hat{V}^2} - \frac{C_0}{T^2} + \frac{bRT}{\hat{V}^3} - \frac{a}{\hat{V}^6} + \frac{c}{\hat{V}^3 T^2} \left(1 + \frac{\gamma}{\hat{V}^2} \right) e^{-\left(\frac{\gamma}{\hat{V}^2} \right)}$$

(54)

where $A_0, B_0, C_0, a, b, c, \alpha$, and γ are all empirical constants for specific substances. The complexity and lack of constant values for an equation of this sort is a weakness of some generalized equation correlations.

The equations given predict vapor behavior to high degrees of accuracy but tend to give poor results near and within the liquid region. The compressibility factor can be used to accurately determine gas volumes when used in conjunction with a virial expansion or an equation such as equation 53 (77). However, the prediction of saturated liquid volume and density requires another technique. A correlation was found in 1958 between the critical compressibility factor and reduced density, based on inert gases. From this correlation an equation for normal and polar substances was developed (78):

$$\log \left(\frac{V_r^{sat}}{\hat{V}_c} \right) = (1 - T_r)^{2.7} \log Z_c$$

(55)

which can predict saturated liquid volumes, V_r^{sat} , to within 1.5%.

Corresponding states have been used in other equations. For example, the Peng-Robinson equation is a modified Redlich-Kwong equation formulated to better correlate vapor-liquid equilibrium (VLE) vapor pressure data. This equation, however, is not useful in reduced form because it is specifically designed to calculate accurate pressure data. Reduced equations generally presuppose knowledge of the pressure.

Three Parameter Models. Most fluids deviate from the predicted corresponding states values. Thus the acentric factor, ω , was introduced to account for asymmetry in molecular structure (79). The acentric factor is defined as the deviation of reduced vapor pressure from 0.1, measured at a reduced temperature of 0.7. In equation form this becomes:

$$\omega = \log P_r - 1.000$$

(56)

where P_r is evaluated at $T_r = 0.7$. This definition was chosen because of the abundance and accuracy of vapor pressure data near the normal boiling point. Acentric factors for various pure compounds are given in many references (3).

The acentric factor, ω , was the third parameter used (20) in an equation based on the second virial coefficient. This equation was further modified and is suitable for reduced temperatures above 0.5.

$$\left(\frac{B P_c}{RT_c} \right) = \left(\frac{B^{(0)} P_c}{RT_c} \right) + \omega \left(\frac{B^{(1)} P_c}{RT_c} \right)$$

(57)

where

$$\begin{aligned} \left(\frac{B^{(0)} P_c}{RT_c} \right) &= 0.1445 - \frac{0.330}{T_r} - \frac{0.1385}{T_r^2} - \frac{0.0121}{T_r^3} - \left(\frac{0.000607}{T_r^8} \right) \\ \left(\frac{B^{(1)} P_c}{RT_c} \right) &= 0.073 + \frac{0.46}{T_r} - \frac{0.50}{T_r^2} - \frac{0.097}{T_r^3} - \frac{0.0073}{T_r^8} \end{aligned}$$

The second virial coefficient is related to the compressibility factor:

$$Z = 1 + B^{(0)} \frac{P_r}{T_r} + \omega B^{(1)} \frac{P_r}{T_r}$$

(58)

Four Parameter Models. Two- and three-parameter theories are only accurate for simple, normal, and some slightly polar fluids. In order to accurately predict polar fluid behavior a fourth parameter is needed (80). The Stiel polarity factor, χ , is one such fourth parameter and follows from the

equation

$$\log P_r = (\log P_r)^{(0)} + \omega (\log P_r)^{(1)} + \chi (\log P_r)^{(2)}$$

(59)

where $(\log P_r)^{(0)}$ and $(\log P_r)^{(1)}$ are the same as those obtained for a normal fluid by equation 57.

To correlate χ to the acentric factor a quadratic Taylor series in terms of the compressibility factor was formulated. This equation is represented as

$$Z = z^{(0)} + \omega z^{(1)} + \chi z^{(2)} + \chi \omega z^{(3)} + \chi^2 z^{(4)}$$

(60)

for $0 < \omega < 0.65$; $-0.06 < \chi < 0.04$ where the values of $z^{(0)}$ and $z^{(1)}$ are the same as those for normal fluids. The values of $z^{(2)}$, $z^{(3)}$, and $z^{(4)}$ have been determined for the gaseous and liquid regions for reduced temperatures from 0.8 to 1.15 and reduced pressures from 0.2 to 6.0 (81). The average percentage error for polar fluids using this method is below 1% compared to over 10% error if the Pitzer correlation is used.

Application to Other Properties.

Extension of Generalized Charts. In 1975, the usefulness of generalized charts was extended upon the publication of extensive tables of residual enthalpy, entropy, and heat capacity (82). This tabular data has also been converted into graphical form (3). The corresponding equations incorporate the acentric factor: *Residual enthalpy*:

$$\left[\frac{H^* - H}{RT_c} \right] = \left[\frac{H^* - H}{RT_c} \right]^{(0)} + \omega \left[\frac{H^* - H}{RT_c} \right]^{(1)}$$

(61)

Residual entropy:

$$\left[\frac{S^* - S}{R} \right] = \left[\frac{S^* - S}{R} \right]^{(0)} + \omega \left[\frac{S^* - S}{R} \right]^{(1)}$$

(62)

Heat capacity:

$$\left[\frac{C_p - C_p^*}{R} \right] = \left[\frac{C_p - C_p^*}{R} \right]^{(0)} + \omega \left[\frac{C_p - C_p^*}{R} \right]^{(1)}$$

(63)

where $*$ denotes an ideal gas value. Combining these equations with ideal-gas property values gives accurate results for any system obeying the corresponding states theory.

Another property which can be represented by generalized charts is fugacity, ϕ . The fugacity of a substance can be regarded as a corrected vapor pressure. At low pressures (below atmospheric) the use of pressure in the place of fugacity leads to little error in calculations. The fugacity coefficient is defined by

$$\phi = \frac{f}{P}$$

(64)

where ϕ is the fugacity coefficient, f is the fugacity, and P is the pressure. In the limit as the pressure approaches zero the fugacity coefficient is unity. If the equation

$$\ln f_{\text{pure } i} = \int_0^P Z_{\text{pure } i} (d \ln P)$$

(65)

where pure i denotes a pure component i and $Z_{\text{pure } i}$ is its compressibility factor, is evaluated by graphical integration and the results are plotted as the fugacity coefficient vs the reduced pressure for constant reduced temperatures, a generalized chart can be constructed.

A generalized chart for the prediction of the fugacity coefficients for 23 compounds using actual PVT data and averaging or smoothing the curves has been constructed (83). The predicted values from the generalized chart have been compared to the experimental values, and agreement within 4% obtained. No limitations were given, but in general the charts should not be used for highly polar substances, quantum fluids, or associating substances.

One of the most versatile and accurate generalized correlations for the prediction of the fugacity coefficient (3) involves a three-parameter generalized correlation which takes advantage of the acentric factor. The correlation breaks the fugacity coefficient into two parts: ϕ^0 and ϕ^1 .

$$\phi = (\phi^0) (\phi^1)^\omega$$

(66)

Generalized charts are available for the calculation of ϕ^0 and ϕ^1 based on experimental data (61,65,66). A generalized chart for methane and pentane has also been made (84).

Generalized Surface Tension Correlations. Use of the principle of corresponding states has provided a practical and accurate method for the estimation of surface tensions. The functional relationship for the surface tension of a pure substance (85) is

$$\sigma_r = \sigma_r(T_r; \alpha_k)$$

(67)

where

$$\sigma_r = \frac{\sigma}{P_c^{2/3} (xT_c)^{1/3}}$$

(68)

x is a constant that depends on the units of P_c and T_c (86) and

$$\sigma = 4.6 \times 10^{-4} \left[P_c^{2/3} T_c^{1/3} \alpha_k (1 - T_r)^{11/9} \right]$$

(69)

$$\alpha_k = 0.1207 \left[1 + \frac{T_b (\ln P_c - 11.526)}{T_c \left(1 - \frac{T_b}{T_c} \right)} \right] - 0.281$$

For 84 simple inorganic substances and a variety of organic compounds, equation 69 produced an average error of 3.0% (87). The above correlations

should not be applied to the quantum substances, associating substances, or highly polar substances.
In 1971 the Stiel polar factor was incorporated in an equation that accounts for polar fluids (88)

$$s = P_c^{2/3} T_c^{1/3} \left(s_r \Big|_{T_r = 0.6} \right) \left(\frac{1 - T_r}{0.4} \right)^m \tag{70}$$

where

$$s_r \Big|_{T_r = 0.6} = 0.1574 + 0.359v - 1.769x - 13.69x^2 - 0.510v^2 + 1.298$$

and

$$m = 1.210 + 0.5385\omega - 14.61\chi - 32.07\chi^2 - 1.65\omega^2 + 22.03\omega\chi$$

This equation was empirically derived from 16 polar fluids and has an average error of 2.9%. A technique for estimating surface tension using nonretarded Hamaker constants (89) has also been presented.

Generalized Correlations for Viscosity. Gas viscosity has also been predicted by corresponding states theory (90) using

$$\eta\xi \left[0.807T_r^{0.618} - 0.357e^{(-0.449T_r)} + 0.340e^{(-4.058T_r)} + 0.018 \right] F_P^0 F_Q^0 \tag{71}$$

where ξ is the reduced, inverse gas viscosity and is defined as

$$\xi = 0.176 \left(\frac{T_c}{M^3 P_c^4} \right)^{1/6} \tag{72}$$

where M is the molecular weight of the gas. For nonpolar, nonquantum gases the values of the low pressure polarity and quantum factors, F_P^0 and F_Q^0 , are equal to 1.

Miscellaneous Generalized Correlations. Generalized charts and corresponding states equations have been published for many other properties in addition to those presented. Most produce accurate results over a wide range of conditions. Some of these properties include: (1) transport properties (64,91); (2) second virial coefficients (80,92); (3) third virial coefficients (72); (4) liquid mixture activity coefficients (93); (5) Henry's constant (94); and (6) diffusivity (95).

Reference Substances. Use of a reference substance has its origins in the work of Clausius-Clapeyron and equation 73, a form of equation 7:

$$\ln P^{\text{sat}} = \frac{\Delta H^{\text{vap}}}{RT} + I \tag{73}$$

where I is an integration constant. If the natural log of the vapor pressure, P^{sat} , is plotted vs the inverse of absolute temperature, T , then the slope of the line becomes the heat of vaporization, ΔH^{vap} , divided by the universal gas constant. A plot called a Cox chart can be constructed by plotting the $\ln P^{\text{sat}}$ of an unknown compound vs the $\ln P^{\text{sat}}$ of a well-documented (3) substance at equal temperatures. By recording only the temperature at which the reference vapor pressure was measured on the x axis, the Cox chart becomes a valuable tool for estimating vapor pressures at temperatures for which no experimental values are available.

In another type of reference plot the temperature of a property of compound A is plotted vs the temperature of the reference substance at equal vapor pressures values.

The values for a single property of two compounds, A and B, are useless unless these values are compared at equal temperature or pressure. Then a deviation from some intermediate value can be determined. If this intermediate value is chosen to be the value of one particular substance, ie, A, the reference substance, both A and B can then be expressed as functions of the reference substance. One very simplistic example is specific gravity where the density of a compound is expressed as the actual density divided by the density of water at 4°C and water is the reference substance.

By use of Othmer plots of reference substances, large tables of thermodynamic data can be expressed as simple correlations which are extremely accurate and easy to use. The real power of these correlations is the ability to interpolate and extrapolate the correlations beyond the experimental values with considerable accuracy. Mathematically stated

$$\log(\text{property}) = m \log(\text{property}_{\text{ref}}) + C \tag{74}$$

where the coefficient m and C are determined by graphing experimental data for the chemical of interest at the same critical difference temperature, T_D , where $T_D = T_c - T$, vs reference values.

Correlations based on reference substances are limited to compounds which have experimentally determined values, but the number of data points needed to produce a correlation is relatively small. A reference substance should be as chemically and physically compatible to the chemical with which it is being compared as possible. Use of reference substances in the ideal assumes no property deviations, thus the smaller the deviations, the lower the absolute error in the correlations.

Values for many properties can be determined using reference substances, including density, surface tension, viscosity, partition coefficient, solubility, diffusion coefficient, vapor pressure, latent heat, critical properties, entropies of vaporization, heats of solution, colligative properties, and activity coefficients. Table 1 lists the equations needed for determining these properties.

Table 1. Othmer Plot Equations

Property	Equation ^a	Comments
surface tension, γ	$\log \gamma = m \log \gamma' + C$	m = ratio of surface activation energies
liquid density, d	$\log d = m \log d' + C$	$m \approx 1$ for most organic liquids
viscosity, η	$\log \eta = m \log P' + C$	P' is the vapor pressure of the reference substance
solubility in liquids, x	$\log x = \frac{(h_l - h_s)}{L'} \log P' + C$	L' is the molal latent heat of reference substance, b_i
		and h are the molal heat content of the solid and saturated solution, respectively
partition coefficients, K_D	$\log K_D = \frac{\Delta H}{L'} \log P' + C$	ΔH is molal heat of transfer of solute between the two layers
diffusion coefficients, D	$\log D = \left(\frac{E_d}{L'} \right) \log P' + C$	E_d is the energy of activation for diffusion

vapor pressure, P

$\log P = m_T \log P' + C'$

m_T is the ratio of the latent heats per mole

^a The superscript prime denotes the reference substance.

Compared to the theory of corresponding states, the reference substance method gives highly accurate results for compounds having sparse experimental data. The corresponding states method gives moderate accuracy for numerous compounds even without actual data.

Empirical Representations of Data

Fitting Simple Functions. In many engineering applications there may be no apparent theoretical basis for the relationship of two variables or the relationship may be too complex to apply. Thus the search for a correlating equation form may at first be along empirical lines. A simple plot of the data in ordinary Cartesian coordinates gives an immediate indication of the essential form of the data.

Each of the graphs shown in Figure 1 suggests an equation given in Table 2 by which a data set may be correlated. For most of these equations, the method of reduction to a linear form may be apparent, possibly through a transformation of variables. Typically, it is desirable to reduce the information to a linear form because the constants of the resulting linear equation can be found either directly from the slope and intercept of the straight line graph, or by means of analytical techniques.

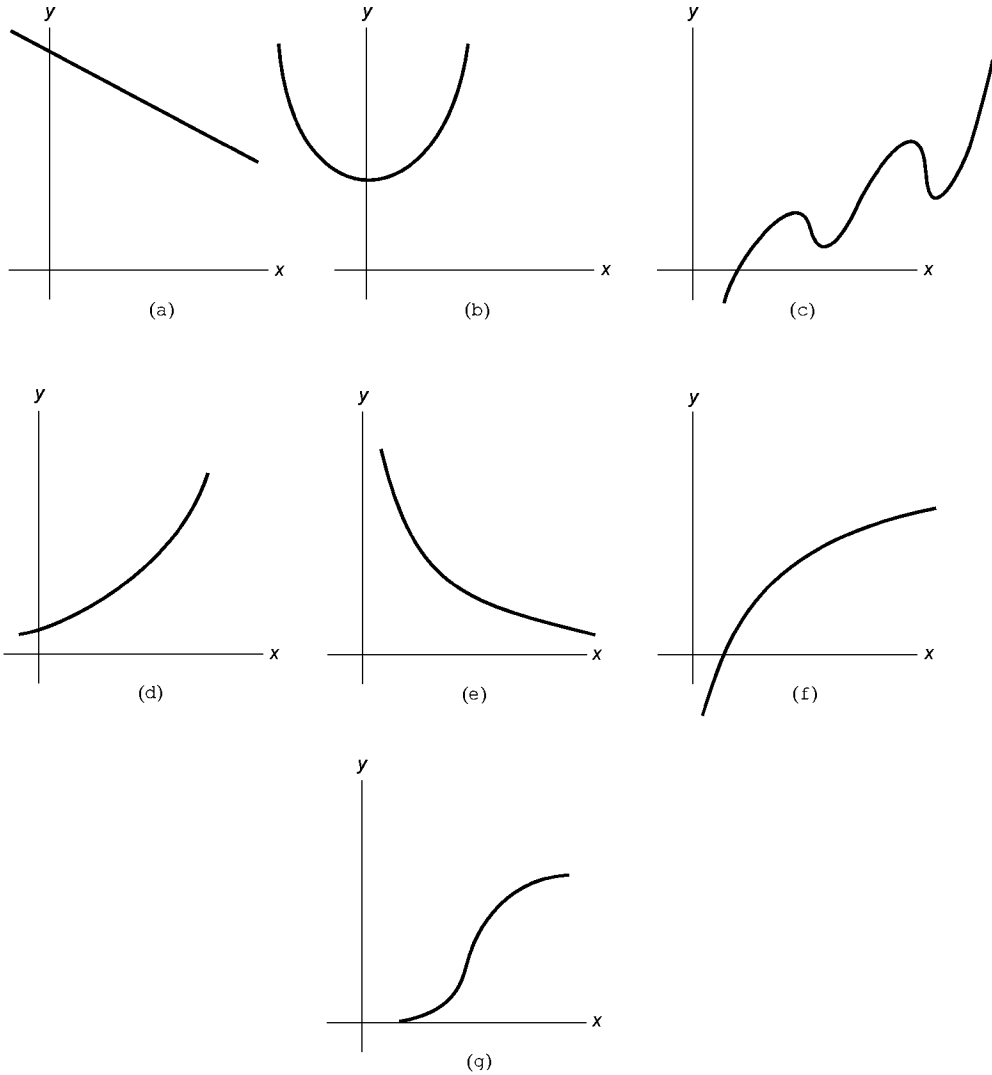


Fig. 1. Some common curves: (a) linear; (b) parabolic; (c) polynomial; (d) exponential; (e) hyperbolic; (f) logarithmic; and (g) logistic. See Table 2 for corresponding equations.

Table 2. Curve Equations and Linear Reduction Methods

Curve	Equation	Method of reduction
linear	$y = a + bx$	plot y vs x ; $a = y$ - intercept; $b =$ slope
parabolic	$y = a + cx^n$	first obtain $a = y$ - intercept of y vs x plot; then plot $\log (y - a)$ vs $\log x$; $c =$ antilog of new y -intercept at $x = 1$; $n =$ slope
polynomial quadratic ^a	$y = a + bx + cx^2$	first obtain $a = y$ - intercept on y vs x plot; then plot $y - y_n - x - x_n$ vs x where y_n, x_n are the coordinates of any point on a smooth curve through the experimental points; $c =$ slope; $b + cx_n = y -$ intercept
exponential	$y = ab^x$ $y = ae^{bx}$	plot $\log y$ vs x ; $a =$ antilog of y -intercept; $b =$ antilog of slope
hyperbolic	$y = a + b/x$	plot $\ln y$ vs x ; $a =$ antilog of y -intercept; $b =$ slope
logarithmic	$y = a \log x$	plot y vs $1/x$; $a = y$ - intercept; $b =$ slope
logistic	$1/y = a + be^{-x}$	plot $1/y$ vs e^x ; $a = y$ - intercept and $b =$ slope or plot $1/y$ vs x ; $a = y$ - intercept; then plot $\log (\frac{1}{y} - a)$ vs $\log x$; $b =$ antilog of y -intercept

^a For the general polynomial quadratic equation, $y = z + bc + cx^2 + dx^3 + \dots$, graphical procedures are almost impossible.

Once the form of the correlation is selected, the values of the constants in the equation must be determined so that the differences between calculated and observed values are within the range of assumed experimental error for the original data. However, when there is some scatter in a plot of the data, the best line that can be drawn representing the data must be determined. If it is assumed that all experimental errors (ϵ) are in the y values and the x values are known exactly, the least-squares technique may be applied. In this method the constants of the best line are those that minimize the sum of the squares of the residuals, ie, the difference, α , between the observed values, y , and the calculated values, Y . In general, this sum of the squares of the residuals, R , is represented by

$$R = \sum_{i=1}^n \alpha_i^2 = \sum_{i=1}^n (y_i - Y_i)^2$$

(75)

For a linear correlation in Y with one independent variable, x , and constants a and b :

$$Y = a + bx$$

(76)

Substitution for Y in equation 75 yields

$$R = \sum_{i=1}^n (y_i - a - bx_i)^2 = 0$$

(77)

where n is the number of data points for which a correlation is required. Values of the two coefficients that minimize R in this equation are found by taking its partial derivative with respect to each of the coefficients and setting these derivatives equal to zero:

$$\frac{\partial R}{\partial a} = 2 \sum_{i=1}^n (y_i - a - bx_i) = 0$$

(78)

$$\frac{\partial R}{\partial b} = 2 \sum_{i=1}^n x_i (y_i - a - bx_i) = 0$$

(79)

Rearrangement yields a set of linear equations with two unknowns:

$$\sum_{i=1}^n a + \sum_{i=1}^n bx_i = \sum_{i=1}^n y_i$$

(80a)

$$\sum_{i=1}^n ax_i + \sum_{i=1}^n bx_i^2 = \sum_{i=1}^n x_i y_i$$

(80b)

Additional rearrangement gives the so-called normal equations:

$$na + b \sum x_i = \sum y_i = 0$$

(81)

$$a \sum x_i + b \sum x_i^2 = \sum x_i y_i = 0$$

(82)

which can be readily solved for a and b as follows:

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

(83)

$$a = \frac{\sum y_i - b \sum x_i}{n}$$

(84)

Thus finding the best values for the slope, b , and intercept, a , of a straight line simply involves developing the sums indicated and combining them in equations 83 and 84.

The least-squares procedure can be applied to the transformed variables of any of the equations in Table 2, where a simple transformation of one or both of the variables results in a linearized expression. The sums for equations 83 and 84 must be formed from the transformed variables rather than from the original data.

Prior to correlation of the data, all that can be considered is the scatter of all the y_i values about the mean $\bar{y}$:

$$\sum (y_i - \bar{y})^2$$

(85)

After obtaining a correlation, it is more meaningful to speak of the scatter of the y_i values around the correlating function:

$$\sum (y_i - Y)^2$$

(86)

Subtracting $(y_i - \bar{y})$ from $(y_i - Y)$ gives $(Y - \bar{y})$, which is that portion of the deviation of any data point from the mean which is explained by the correlation. The coefficient of determination, γ^2 , a statistical parameter which varies from 0.0 to 1.0, is defined as:

$$\gamma^2 = \frac{\sum (Y - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

(87)

How well the developed equation fits the data is given by the correlation coefficient, γ (0.0 = poor fit, ± 1.0 = excellent fit), which is the square root of the

coefficient of determination. The physical significance of this ratio is apparent from the graph shown in Figure 2. Simple algebraic transformation gives the following more useful form:

$$\gamma^2 = \frac{\left[\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n} \right]^2}{\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n} \right] \left[\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n} \right]} \tag{88}$$

This involves the same summations of data values which were required in the evaluation of the constants a and b and can be determined at the same time.

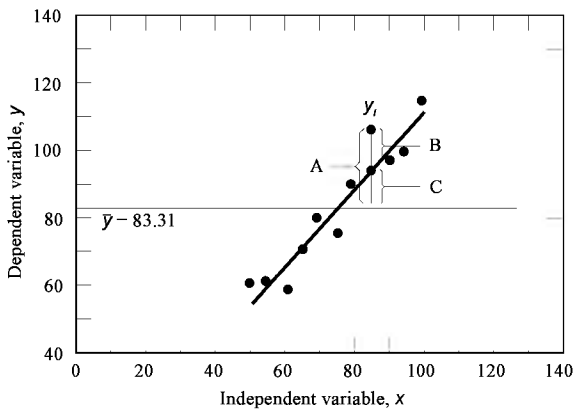


Fig. 2. Least-squares plot showing the determinants for the coefficient of determination. A, total deviation ($y_i - \bar{y}$); B, unexplained deviation ($y_i - y$); and C, explained deviation ($y - \bar{y}$).

The least-squares technique can be extended to any number of variables as long as the equation is linear in its coefficients. The linear correlation of y vs x can be extended to the correlation of y vs multiple independent variables generating an equation of the form:

$$Y = a + bx_1 + cx_2 + dx_3 + \cdots \Omega x_n \tag{89}$$

where $x_1, x_2, \dots, x_n$ are n different variables. For example, y might be the yield of a reaction product, whereas x_1 might be the reaction temperature, x_2 the reaction pressure, x_3 the concentration of a catalyst, x_4 the space velocity, and so on.

The coefficients of equation 89 can be determined by an extension of the procedure outlined earlier, ie, by minimizing the sum of the squares of the deviations of observed y values from those that would be predicted by the model. Substitution of equation 89 into equation 75 yields:

$$R = \sum_{i=1}^n (y_i - a - bx_{1i} - cx_{2i} - dx_{3i} - \cdots \Omega x_{ni})^2 \tag{90}$$

Setting

$$\frac{\partial R}{\partial a} = 0, \quad \frac{\partial R}{\partial b} = 0, \quad \text{etc} \tag{91}$$

results in a set of simultaneous equations that determine simple sums of terms involving the experimentally determined data values.

$$ma + b\Sigma x_1 + c\Sigma x_2 + d\Sigma x_3 + \cdots \Omega \Sigma x_n = \Sigma y \tag{92}$$

$$a\Sigma x_1 + b\Sigma x_1^2 + c\Sigma x_1 x_2 + d\Sigma x_1 x_3 + \cdots \Omega \Sigma x_1 x_n = \Sigma x_1 y \tag{93}$$

$$a\Sigma x_2 + b\Sigma x_1 x_2 + c\Sigma x_2^2 + d\Sigma x_2 x_3 + \cdots \Omega \Sigma x_2 x_n = \Sigma x_2 y \tag{94}$$

$$a\Sigma x_n + b\Sigma x_1 x_n + c\Sigma x_2 x_n + d\Sigma x_3 x_n + \cdots \Omega \Sigma x_n^2 = \Sigma x_n y \tag{95}$$

In principle, this set of equations can be solved for the various constants, a through Ω , just as a and b were obtained previously. In practice, however, the actual numerical evaluation involves considerable computation in all but the simplest examples. Computer solution by matrix techniques designed specifically to handle this type of data correlation problem is usually required.

Curve fitting to data is most successful when the form of the equation used is based on a known theoretical relationship between the variables associated with the data points, eg, use of the Clausius-Clapeyron equation for vapor pressure. In the absence of known theoretical relationships, polynomials are one of the most useful forms to describe a curve. Polynomials are easy to evaluate; the coefficients are linear; and the degree, ie, the highest power appearing in the equation, is a convenient measure of smoothness. Lower orders yield smoother fits.

Fitting polynomials to data points involves essentially the same technique as for correlation of multiple variables. Replacing x_1 through x_n in equation 89 by x^1 through x^n results in the polynomial:

$$Y = a + bx + cx^2 + dx^3 + \cdots \Omega x^n \tag{96}$$

Continuing this substitution through the resulting set of simultaneous equations and solving them generates the coefficients of this polynomial equation. This process is termed polynomial regression or curvilinear regression. It should be noted that equation 96 is still linear in the coefficients a through Ω .

A nonlinear model would involve one or more of the coefficients in nonlinear form such as the Antoine equation (eq. 8) or

$$Y = a + be^{-cx} \tag{97}$$

or

$$Y = ae^{bx} + cx^2$$

(98)

The nonlinear constant c in these equations cannot be evaluated directly by the methods previously described. Even forms such as these can be handled, however. For example, subtracting a trial value of a from y and taking logarithms transforms equation 97 into the linear form:

$$\ln(Y - a) = \ln b + cx$$

(99)

from which the constants b and c as well as the correlation coefficient can be determined. Iterations on the trial values of the constant a establishes the value which results in the best fit. Following well-known techniques, the familiar Antoine vapor pressure correlation (eq. 8) can be evaluated using the substitutions $a = A$; $b = AC - B$; and $c = -C$; yielding

$$aT + b - c\log_{10} P^{\text{sat}} = T\log_{10} P^{\text{sat}} - 0$$

(100)

which can be handled quite easily by standard techniques. Upon substitution of sufficient data to yield the constants a , b , and c one can find the Antoine constants from $A = a$; $B = -(ac + b)$; and $C = -c$. Equation 100 can represent the vapor pressure of many pure substances from the triple point to the critical point.

Some formulas, such as equation 98 or the van der Waals equation, are not readily linearized. In these cases a nonlinear regression technique, usually computational in nature, must be applied. For such nonlinear equations it is necessary to use an iterative or trial-and-error computational procedure to obtain roots to the set of resultant equations (96). Most of these techniques are well developed and include methods such as successive substitution (97,98), variations of Newton's rule (99–101), and continuation methods (96,102).

The usual practice in these applications is to concentrate on model development and computation rather than on statistical aspects. In general, nonlinear regression should be applied only to problems in which there is a well-defined, clear association between the independent and dependent variables. The generalization of statistics to the associated confidence intervals for nonlinear coefficients is not well developed.

Nomographs. The word *nomograph*, from the Greek, means the image of the law, so literally a nomograph is any diagram that represents a mathematical function. The function is evaluated by drawing lines through axes. These lines intersect a result curve. There are three types of nomographs, the Cartesian graph, slide rule, and alignment chart.

Alignment charts, developed in 1899 (103), are sometimes difficult to interpret. Whereas the Cartesian graph of a line looks like a line, the alignment diagram of a line is a point between two axes. By extension, a three-axis Cartesian surface may be represented by a line straddling three axes. This ability to condense an immense amount of data to simple graphical correlations is the strength of alignment charts and nomographs.

Nomographs were first introduced in the form of the Cartesian graph. Characterized by perpendicular axes and reference lines, two-axis Cartesian graphs can be designed to show the solution to any two-variable equation. Expression of three-variable equations is possible with the use of families of curves or three-axis surfaces. The former involves considerable interpolation, because a desired value often lies somewhere between two curves on the graph. Unlike a Cartesian graph, the axes of a nomograph are not necessarily perpendicular, nor parallel, nor even linear. The main feature of an alignment chart is the reference line, which is not necessarily perpendicular to any axis. All nomography is based on geometrical construction and there are great differences in the appearance of various forms of nomographs.

Comparison of Alignment Charts and Cartesian Graphs. There are typically fewer lines on an alignment chart as compared to Cartesian plots. This reduces error introduced by interpolation and inconsistency between scales. For example, to find a point (x,y) on a Cartesian graph one draws two lines, one perpendicular to each axis, and these reference lines intersect at the point (x,y) . This point (x,y) may correspond to some finite value found by reading a contour map represented by a family of curves corresponding to different values of the function.

An alignment chart is used by drawing one reference line through the two axes. This reference line, which need not be perpendicular to either axis, crosses a result axis at a unique finite value. This result axis represents the contour map on a Cartesian graph. Thus each line on an alignment chart represents a point on a Cartesian graph.

Construction of Alignment Charts. Of the ways to construct alignment charts, the brute force method, which requires some idea of the geometry for the chart, is the easiest method to use. The mathematical method, which uses parametric equations of scale to determine the placement and scale of each axis, is the most accurate, but the most difficult to apply.

Brute Force Method. An example of this method is demonstrated in Figure 3 for the three-variable equation, $x + 3y = z$, which requires two known points in order to find the third. The first step is to determine two axes and the appropriate scales, drawing two parallel lines and marking the lines with evenly spaced graduations as shown in Figure 3. This is the basic form of the parallel alignment chart. The scales should be evenly spaced. The third axis is determined by evaluating the equation at several points. The position of the third axis is found by evaluating the equation for several sets of points which share the same value for the third axis. This can be done for two values, in order to find two points which lie on the unknown axis. A straight line is used to define the axis. The scales are marked similarly by evaluating the equation for useful values on the third axis. Because any value of z is found only on one point on each of the other axes, the placement of the z -axis can be determined by choosing an arbitrary value for z and solving for values of x and y which satisfy the equation. For example, if $z = 13$, then if $y = 1$, then $x = 13 - 3(1) = 10$, giving $(x, y, z) = (10, 1, 13)$; if $y = 3$, then $x = 13 - 3(3) = 4$, giving $(x, y, z) = (4, 3, 13)$. The lines drawn connecting each of these pairs of points intersect at the point on the z -axis, where $z = 13$. Another point on the axis can be found by choosing a new value for z . Setting $z = 40$, if $y = 10$, then $x = 40 - 3(10) = 10$, giving $(x, y, z) = (10, 10, 40)$; if $y = 9$, then $x = 40 - 3(9) = 13$, giving $(x, y, z) = (13, 9, 40)$. A vertical line connects these intersection points to define the z -axis.

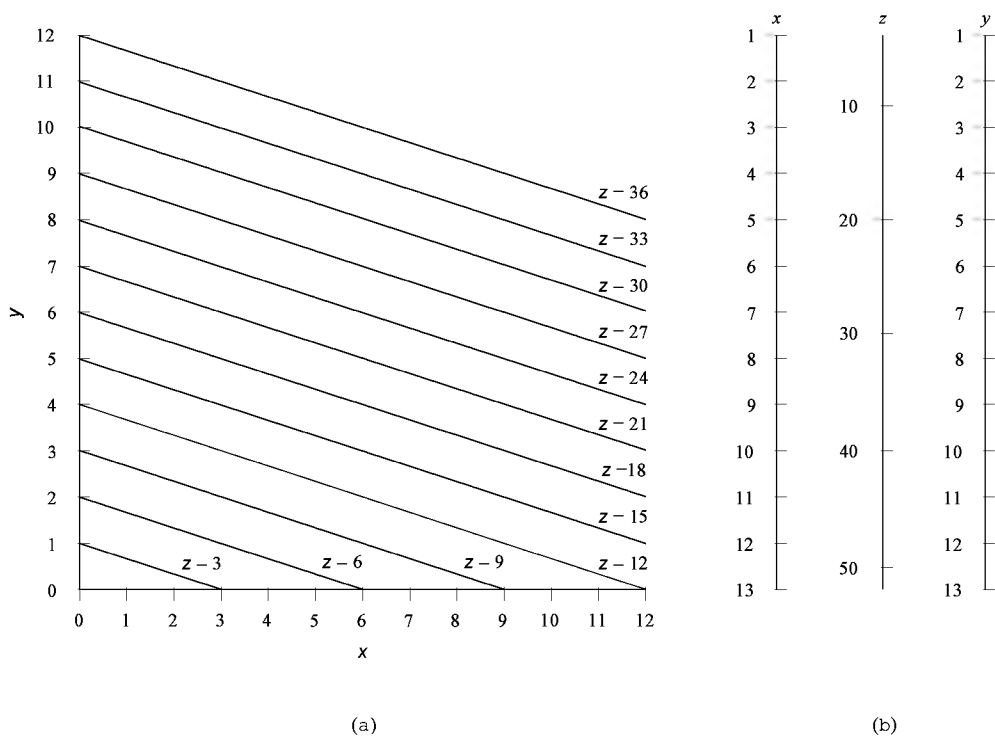


Fig. 3. Comparison for the equation $x + 3y = z$ of (a) Cartesian and (b) nomographic formats. See text.

The scale can be determined in much the same way, by choosing the values to mark on the axis and evaluating the equation for sets of points. The limits of this axis can be found by evaluating the extremes. For $x = 1, y = 1, z = 4$; for $x = 13, y = 13, z = 65$. So values of z can be chosen anywhere between 4 and 65. If increments of 10 are desired, then values of x and y which, when evaluated with the equation, give the desired values of z incremented by 10, should be found. A comparison between the Cartesian representation and the alignment chart equivalent is shown in Figure 3. The alignment chart is simpler because of the smaller number of lines.

Mathematical Derivation of Scales. The brute force method for determining scale is not always the most convenient. Scale direction and axis placement can be difficult to guess. In addition, there may be some advantage to nonlinear axes. An accurate method, based on the use of parametric equations for determining the scale has been developed using determinants (104–107). For complex equations, it is possible to break the problem into several parts that can be simultaneously evaluated. Because any function can be broken into a set of parametric functions, the task is to graphically interpret the functions. The parametric equations can be converted into equations of scale which differ from chart to chart, depending on whether the function involved is linear (eq. 101x), exponential (eq. 101y), or inverse (eq. 101z):

$$L[t] = mf(t) + n \tag{101x}$$

$$L[t] = m\log f(t) + n \tag{101y}$$

$$L[t] = mf(t)^{-1} + n \tag{101z}$$

where L is the linear placement of the scale value; $f(t)$ is the value of the function; t is to be marked on the scale; m is the modulus of the scale and determines the length of the axis; and n determines the placement of the scale. These parametric equations of scale can be expanded to be used on a Cartesian plane with both x - and y -axis coordinates

$$L[t] = \{m_x f(t) + n_x, m_y f(t) + n_y\} \tag{102}$$

giving an x - and a y -coordinate for each mark of the t -axis. A full axis can be derived with this equation. A set of parametric equations can be derived that produce simultaneous solutions to a function. The challenge is to align them properly.

Each of the parametric equations that can be formed from an expression represents an axis on the chart. Each set of parametric equations must simultaneously agree with the equation they represent. In other words, on a line drawn through any two variables, a third variable can be found which satisfies the parametric equations. So an evaluation of the chart requires that values produced by each parametric equation be on the chart as a line. A determinant can be used to determine whether or not points are collinear. The parametric equations must be evaluated so they always produce values which lie on a line. By replacing the x and y points with parametric equations of scale for the chart, it is possible to create any diagram. This method can be used to determine the placement of the axes, because the parametric equations can be transformed into equations of scale.

Various Nomographs. The alignment chart is restricted neither to addition operations, nor to three-variable problems. Alignment charts can be used to solve most mathematical problems, from linear ones having any number of variables, to ratiometric, exponential, or any combination of problems. A very useful property of these alignment diagrams is the fact that they can be combined to evaluate a more complex formula. Nomographs for complex arithmetical expressions have been developed (108).

Computer Techniques. Rapid development of a wide range of data regression techniques has occurred in the latter 1900s. Basic to most engineering work stations are databases (qv) for efficient data retrieval and management, and spreadsheets (109) in which numerous routine calculations can be quickly correlated, regressed, or otherwise manipulated (see also COMPUTER AIDED DESIGN AND MANUFACTURE (CAD/CAM); COMPUTER AIDED ENGINEERING (CAE)).

The ability to perform complex regressions and matrix inversions are subjects in numerous undergraduate and graduate courses (110). Problems requiring estimated solutions to sparse matrices, eg, group contribution parameter estimations, have been found to be mathematically similar to annealing of metals and mathematical Boltzmann machines (111,112). Numerous software packages have been developed for specific applications, such as kinetics (113) and chemical formulations (114), and as general tools for modeling and reformulation (115) and expert systems (qv) (116).

Molecular-Based Methods

In the most general sense, a molecular-based property prediction method can be thought of as any method in which the physical features of a molecule (or

molecules for multicomponent properties) are correlated to a property. Some methods, such as molecular thermodynamics, generalize characteristics of whole molecules to correlate them to bulk properties consistent with classical thermodynamics. Other methods correlate molecular fragments, bonds, atomic compositions, molecular topological features, and solution interactions.

Molecular Thermodynamics, the Kinetic Theory of Gases, and Statistical Thermodynamics. Molecular thermodynamics attempts to explain observed physical properties via individual molecular properties (117). It uses classical thermodynamics, as well as concepts from statistical thermodynamics and chemical physics (118). The molecular view of thermodynamics can be traced back to 1738, when Bernoulli proposed that a gas consisted of individual molecules in constant motion, colliding with each other and with their containment vessels. During the nineteenth century, Maxwell and Boltzmann developed this into the kinetic theory of gases (119,120,121), which applies the laws of mechanics to microsystems. From this theory expressions for the pressure of a gas, its internal energy, and its specific heat capacity can be derived (122). Statistical thermodynamics, or equilibrium statistical mechanics, provides the mathematical glue between the kinetic theory and macrosystems, ie, systems containing large numbers of molecules. It takes advantage of the fact that molecules are very numerous and that average properties can be estimated through probability and statistical analysis (122).

For example, the measured pressure exerted by an enclosed gas can be thought of as a time-averaged manifestation of the individual molecules' random motions. When one considers an individual molecule, however, statistical thermodynamics would propose its random motion or pressure could be quite different from that measured by even the most sensitive gauge which acts to average a distribution of individual molecule pressures. The particulate nature of matter is fundamental to statistical thermodynamics as opposed to classical thermodynamics, which assumes matter is continuous. Further, these elementary particles and their complex substructures exhibit wave properties even though intra- and interparticle energy transfers are quantized, ie, not continuous. Statistical thermodynamics holds that the impression of continuity of properties, and even the solidity of matter is an effect of scale.

The kinetic theory attempts to describe the individual molecules' energies and interactions; statistical thermodynamics attempts to fundamentally develop the equation of state from considerations of groupings of molecules. These approaches are complementary in many ways (3,123,124). A well-referenced text covering molecular thermodynamics is also available (125).

Basic Problem in Statistical Thermodynamics and Quantum States. All problems in statistical thermodynamics have been postulated to be essentially forms of the same problem (126), that is, given a constant energy level, E , for a closed set of identical systems, eg, molecules, determine the distribution of these systems over all possible states in which the set can exist. This set is often referred to as an assembly or ensemble. Implicit in this description is that whereas there is interaction between the systems, the interaction is so weak that it can be disregarded. Thus, from a mathematical point of view, each system is thought of as containing its own distinct energy and E is obtained by summing all distinct system energies.

Distinct molecular energies are found at discrete, or quantized, levels. One method to determine the energy associated with a quantum state of a system is by describing the spins, or angular-momentums of the components and subcomponents. The wave functions which are used in these descriptions, eg, Schrödinger's wave equation (126,127), may be symmetric or antisymmetric depending on the spins. An antisymmetric wave function is one in which the sign of the wave function is dependent on component positions within the system. Components having even numbers of subcomponents have integral spins, and therefore symmetric wave functions, ie, sign not dependent on component position. The Pauli exclusion principle (128) indicates that within a system of components having antisymmetric wave functions, no two particles can have the same quantum state.

Systems containing symmetric wave function components are called Bose-Einstein systems (129); those having antisymmetric wave functions are called Fermi-Dirac systems (130,131). Systems in which all components are at a single quantum state are called Maxwell-Boltzmann systems (122). Further, a boson is a particle obeying Bose-Einstein statistics, a fermion is one obeying Fermi-Dirac statistics (132).

Maxwell-Boltzmann particles are distinguishable, and a partition function, or distribution, of these particles can be derived from classical considerations. Real systems exist in which individual particles are indistinguishable. For example, individual electrons in a solid metal do not maintain positional proximity to specific atoms. These electrons obey Fermi-Dirac statistics (133). In contrast, the quantum effects observed for most normal gases can be correlated with Bose-Einstein statistics (117). The approach to statistical thermodynamics described thus far is referred to as wave mechanics. An equivalent quantum theory is referred to as matrix mechanics (134–136).

Ensembles and Postulates. Once estimations of time-averaged energies for specific quantum states of identical systems have been made, a large number of identical systems called an ensemble is defined. An ensemble is typically defined according to a set of boundary conditions encountered in a real system. The ensemble is a mathematical construct that facilitates summation of all possible states. The typical method is to assign a state to each system in an ensemble. If the number of systems is great enough, summing an ensemble of systems with fixed states arrives at the same total as averaging all possible states of a single system and multiplying by the number of systems. The microcanonical and canonical ensembles are the two most frequently encountered in chemical thermodynamics (118). A microcanonical ensemble represents an isolated system having independent parameters, E ; volume, V ; and N molecules. A canonical ensemble represents a closed isothermal system with independent parameters N , V , and temperature T . An ensemble encountering fewer mathematical difficulties is the grand canonical ensemble in which restrictions on E and N are relaxed. T , V , and the chemical potentials $\mu_1, \mu_2, \dots$ of components 1,2, ... are the usual independent parameters.

Mechanical–thermodynamic property values (137) for each system are assigned based on quantum theory and the following postulates: (1) the most probable distribution of systems among all possible quantum states is analogous to the equilibrium property value from classical thermodynamics, and (2) the most probable distribution of systems among all possible quantum states is that distribution for which the number of quantum states of the systems and their surroundings is a maximum. Mechanical–thermodynamic properties are those for which a system in a quantum state would have a well-defined value, eg, pressure, volume, energy, and mass. Nonmechanical–thermodynamic properties, eg, temperature and entropy, do not have well-defined values for a system, and therefore can only be evaluated through a comparison with classical thermodynamic relationships. For example, the mass of an ensemble may be found by adding the masses of the constituent systems. However, the same cannot be said of temperature. The term classical refers to property values derived from macroscopic measurements, eg, mass, pressure, volume, and temperature, or calculated from these derived values, eg, enthalpy, entropy, and internal energy. A concise review of the mathematics necessary to calculate properties from these postulates based on Maxwell-Boltzmann particles is available (118).

Applications. Statistical thermodynamics holds great promise as a means to characterize molecules, bonds, reactions, and energy fields in ways which are consistent with both observed and calculated classical thermodynamic properties. However, it is exceedingly difficult, if not impossible, to solve analytically for the trajectories of just three mutually interacting bodies from quantum or classical mechanics (118), and the problem at hand involves on the order of 10^{23} bodies. Few physical property prediction methods are available that are based solely on statistical thermodynamics. However, its methodology is fundamental to group contribution and most other molecular-based physical property prediction methods (122,137–141). Some insight into the applications of statistical thermodynamics to nonequilibrium problems is available (142,143). An alternative approach in which field theory is applied to replace volume and pressure with strain and stress tensors, respectively, involves continuum mechanics (144). Areas in which molecular thermodynamics goes beyond classical thermodynamics is of interest. The second and third coefficients of the virial equation-of-state for real gases have been shown to correspond to binary and ternary molecular interactions through statistical thermodynamics (118,121). From this information, and a basic understanding of intermolecular forces, eg, polar attractions, dipole and quadrupole moments, and dispersion forces (118), these coefficients have been predicted for simple gases (118,138,145). Equations-of-state for a series of low to medium pressure interacting gases have been developed, as well as thermodynamic properties of crystalline solids (138). Spectroscopic properties of various energy radiations related to electron and atomic transitions have been explained largely through statistical thermodynamics (146–148). Mathematical descriptions of radiation processes including gamma and x-rays, radiant heating, the propagation of light and the laser effect are demonstrated through statistical considerations (138,149,150). Other equilibrium statistical thermodynamic demonstrations can be found in prediction of magnetic properties (138,151,152). Nonequilibrium statistical thermodynamics are used to explain and predict Brownian motion (153,154) and other physical phenomena based on particle fluctuations (138,142,143). *Ab initio* techniques are beginning to be developed to directly correlate atomic wave functions to physical properties (155–158); however, this area does not extend far beyond descriptions of simple species such as methane.

Group Contribution Methods. It has been shown that many macroscopic physical properties, ie, those derived from experimental measurements of bulk solutions or substances, can be related to specific constituents of individual molecules. These constituents, or functional groups, are usually composed of commonly found combinations of atoms. One procedure for correlating functional groups to a property is as follows. (1) A set of

substances is selected for which the property has been determined experimentally. (2) Groups are determined for each substance based on a previously determined set of rules. (3) Property contributions for each group are determined based on a previously determined summation equation. This is done with a variety of computer optimization algorithms. (4) Property values are calculated based on the group contributions and compared to the experimental results. As a further test, property values are sometimes estimated for a selection of substances not included in the optimization.

Group contribution methods routinely predict properties of substances, both real and imagined, for which the only information known is structural. For example, if groups are selected such that each group has one carbon atom and its associated hydrogen atoms, then the groups CH₂ and CH₃ would be adequate to describe almost all normal alkanes, methane being the exception. Regression of a given physical property by an appropriate equation using these two groups would allow the prediction of the property of any normal alkane for which the structure could be written.

Aside from the rather trivial matter of predicting molecular weight by summing the mass contribution of constituent atoms, group contribution methods have numerous limitations. (1) Predictions for molecules of drastically different sizes than those in the optimization set usually exhibit large errors. (2) Short of having an experimental determination of the property for the compound of interest, there is no rigorous method for determining the accuracy of a group contribution-based prediction. Without experimental data, prediction of accuracy is inherently extrapolative and there is no means to validate that the compound of interest falls inside the range of application of the correlating equation. (3) Although attempts have been extensive, group contribution methods are predominantly empirical and involve numerous assumptions. Researchers are far from understanding the behavior of groups in a general way, such as the understanding of PVT relationships afforded by classical thermodynamics. (4) Seemingly similar substances are sometimes observed to be exceptions to the rule for many group contribution methods, (5) Percentage error in group contribution methodology can be as much as an order of magnitude or greater than that deemed acceptable for many applications, such as industrial plant design.

Recognizing that there is presently a need for property values for tens of thousands of substances, but experimental data for only a small percentage of these substances, group contribution methods are viewed as the only choices for many problems such as newly or yet-to-be-synthesized compounds, situations where available data are well outside the conditions of interest, and reaction kinetics studies involving unknown intermediates.

A relatively simple example of a group contribution technique is the method for estimating liquid and solid heat capacities (159). This method is a modification of Kopp's rule (160,161) which was originally proposed in 1864. Kopp's rule states that, at room temperature, the heat capacity of a solid compound is approximately equal to a stoichiometric summation of the heat capacities of its atoms (elements). The Hurst-Harrison modified equation is as follows:

$$C_p = \sum_{i=1}^n (C_i \times N_i) + C_{\text{misc}} \times N_{\text{misc}} \tag{103}$$

where C_p is specific heat in J (K·mol) at 298 K; C_i is a constant associated with element i ; N_i is the number of occurrences of element i ; C_{misc} is a constant associated with elements not having a specific C_i ; N_{misc} is the number of occurrences of elements not having a specific C_i ; n is the number of different elements in the compound for which there are specific constants.

Values for 32 elements and the C_{misc} constant have been tabulated (159). An average absolute error of 9.6° for 721 solid compounds (13° standard deviation) and 8.8° for 477 liquid compounds (11.5° standard deviation) has been reported.

QSAR Methods for Fluid Solubility Prediction. Several group contribution methods for predicting liquid solubilities have been developed. These methods as well as other similar methods are often called quantitative structure-activity relationships (QSARs). This field is experiencing rapid development.

UNIFAC and ASOG Development. Pertinent equations of the UNQUAC functional-group activity coefficient (UNIFAC) model for prediction of activity coefficients including example calculations are available (162). Much of the background of UNIFAC involves another QSAR technique, the analytical solution of groups (ASOG) method (163).

Predecessors to QSAR correlations include the Cox chart families (164), the local composition concept (165), and the development of mixing rules and solution theory by many researchers (118). During the 1960s a set of group selection rules was developed (25,32,49) for a broad range of organic compounds and a molecular solution model. Combination of this set of rules, the model, and emerging numerical computational capabilities resulted in ASOG, which allowed for the prediction of activity coefficients using no species-specific data beyond structure. With each successive modification or refinement, modeling of the interactions between molecules in solution improved.

The nonrandom, two-liquid (NRTL) equation (166) was developed along similar lines, and in 1975 the universal quasichemical (UNIQUAC) equation and UNIFAC models were published (28,35). The latter was a group contribution method which was claimed to have comparable accuracy in activity-coefficient estimation to ASOG for vapor–liquid systems. This greatly improved accuracy for liquid–liquid systems. A more rigorous mathematical description of the original method (167) and several revisions have been published (28,35–39,168). Available groups and interactions through the 1991 revision (168) are provided in Table 3 and Figure 4, respectively. In addition, application-specific modifications have been proposed. ASOG development has continued at a slower pace.

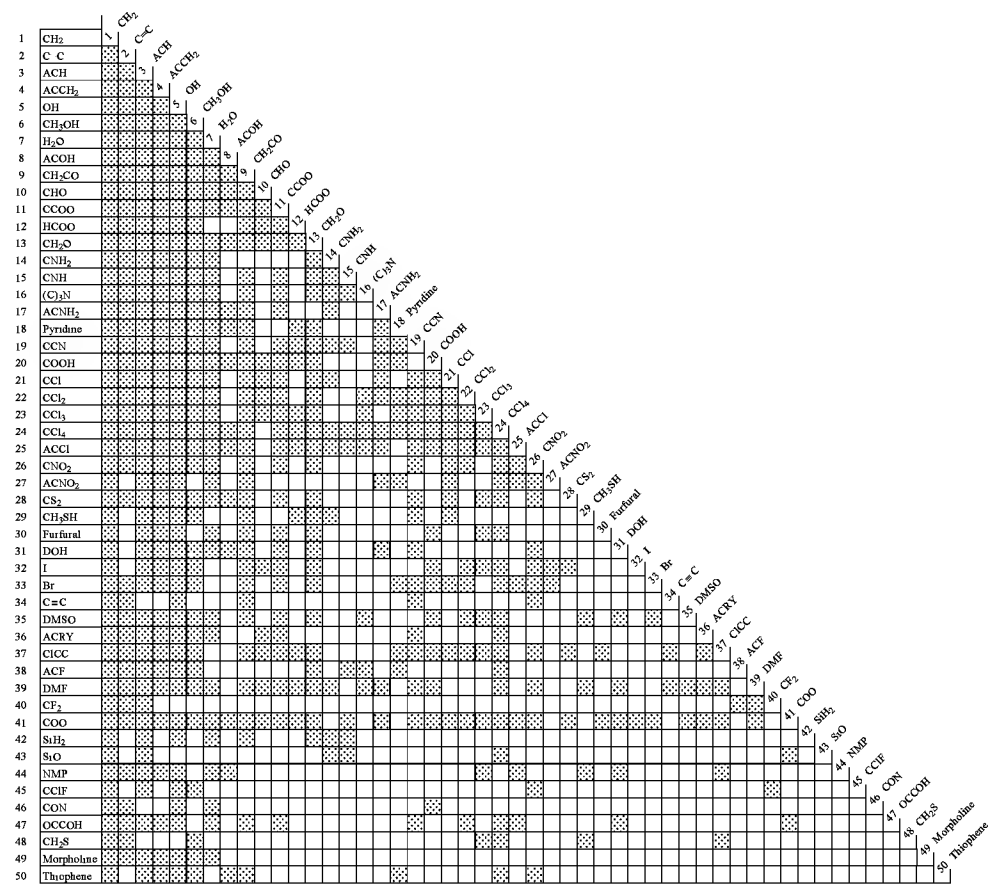


Fig. 4. UNIFAC group interaction parameter matrix, the  represents parameters fit and  parameters not available (168). A represents an aromatic moiety. See Table 3 for additional terms.

Table 3. UNIFAC Functional Groups^{a, b}

Main group	Subgroup	Component	Constituent groups
C—H	C	2,2-dimethylbutane	4 CH ₃ , 1 CH ₂ , 1 C
	CH	2-methylpropane	3 CH ₃ , 1 CH
	CH ₂	pentane	2 CH ₃ , 3 CH ₂
	CH ₃		
	CH ₄	methane	1 CH ₄
CH=CH	C=C	2,3-dimethyl-2-butene	4 CH ₃ , 1 C=C
	CH=C	2-methyl-2-butene	3 CH ₃ , 1 CH=C
	CH ₂ =C	2-methyl-1-butene	2 CH ₃ , 1 CH ₂ , 1 CH ₂ =C
	CH=CH	2-pentene	2 CH ₃ , 1 CH ₂ , 1 CH=CH
	CH ₂ =CH	1-butene	1 CH ₃ , 1 CH ₂ , 1 CH ₂ =CH
ACH	CH ₂ =CH ₂	ethylene	1 CH ₂ =CH ₂
	AC	styrene	1 CH ₂ =CH, 5 ACH, 1 AC
	ACH	benzene	6 ACH
	ACCH	cumene	2 CH ₃ , 5 ACH, 1 ACCH
	ACCH ₂	<i>n</i> -propylbenzene	
ACCH	ACCH ₃	toluene	1 CH ₃ , 1 CH ₂ , 5 ACH, 1 ACCH ₂ 5 ACH, 1 ACCH ₃
OH	OH	2-butanol	2 CH ₃ , 1 CH ₂ , 1 CH, 1 OH
CH ₃ OH		methanol	1 CH ₃ OH
H ₂ O		water	1 H ₂ O
ACOH	ACOH	phenol	5 ACH, 1 ACOH
CH _n CO	CH ₂ CO	(ketone group as any group except second carbon)	
		3-hexanone	2 CH ₃ , 2 CH ₂ , 1 CH ₂ CO
	CH ₃ CO	(ketone group as second carbon)	
		2-pentanone	1 CH ₃ , 2 CH ₂ , 1 CH ₃ CO
CHO	CHO	acetaldehyde	1 CH ₃ , 1 CHO
CH _n COO	CH ₂ COO	pentyl propanoate	2 CH ₃ , 4CH ₂ , 1 CH ₂ COO
	CH ₃ COO	butyl acetate	1 CH ₃ , 3CH ₂ , 1 CH ₃ COO
HCOO	HCOO	methyl formate	1 CH ₃ , 1 HCOO
CH _n O	FCH ₂ O	tetrahydrofuran	3 CH ₂ , 1 FCH ₂ O
	CHO	diisopropyl ether	4 CH ₃ , 1 CH, 1 CHO
	CH ₂ O	diethyl ether	2 CH ₃ , 1 CH ₂ , 1 CH ₂ O
	CH ₃ O	dimethyl ether	1 CH ₃ , 1 CH ₃ O

CH_nNH_2	CHNH_2	isobutylamine	$2\text{CH}_3, 1\text{CH}_2, 1\text{CHNH}$
	CH_2NH_2	propylamine	$1\text{CH}_3, 1\text{CH}_2, 1\text{CH}_2\text{NH}_2$
	CH_3NH_2	methylamine	$1\text{CH}_3\text{NH}_2$
CH_nNH	CHNH	diisopropylamine	$4\text{CH}_3, 1\text{CH}, 1\text{CHNH}$
	CH_2NH	diethylamine	$2\text{CH}_3, 1\text{CH}_2, 1\text{CH}_2\text{NH}$
	CH_3NH	methylethylamine	$1\text{CH}_3, 1\text{CH}_2, 1\text{CH}_3\text{NH}$
CH_nN	CH_2N	triethylamine	$3\text{CH}_3, 2\text{CH}_2, 1\text{CH}_2\text{N}$
	CH_3N	trimethylamine	$1\text{CH}_3, 1\text{CH}_3\text{NH}$
ACNH_2	ACHN_2	aniline	$5\text{ACH}, 1\text{ACNH}_2$
$\text{C}_5\text{H}_n\text{N}$	$\text{C}_5\text{H}_3\text{N}$	2,3-dimethylpyridine	$2\text{CH}_3, 1\text{C}_5\text{H}_3\text{N}$
	$\text{C}_5\text{H}_4\text{N}$	3-methylpyridine	$1\text{CH}_3, 1\text{C}_5\text{H}_4\text{N}$
	$\text{C}_5\text{H}_5\text{N}$	pyridine	$1\text{C}_5\text{H}_5\text{N}$
CH_nCN	CH_2CN	propionitrile	$1\text{CH}_3, 1\text{CH}_2\text{CN}$
	CH_3CN	acetonitrile	$1\text{CH}_3\text{CN}$
CH_nOOH	HCOOH	formic acid	1HCOOH
	COOH	acetic acid	$1\text{CH}_3, 1\text{COOH}$
CH_nCl	CCl	2-chloro-2-methylpropane	$3\text{CH}_3, 1\text{CCl}$
	CHCl	2-chlorobutane	$2\text{CH}_3, 1\text{CH}_2, 1\text{CHCl}$
	CH_2Cl	1-chlorobutane	$1\text{CH}_3, 2\text{CH}_2, 1\text{CH}_2\text{Cl}$
CH_nCl_2	CCl_2	2,2-dichloropropane	$2\text{CH}_3, 1\text{CCl}_2$
	CHCl_2	1,1-dichloropropane	$1\text{CH}_3, 1\text{CH}_2, 1\text{CHCl}_2$
	CH_2Cl_2	dichloromethane	$1\text{CH}_2\text{Cl}_2$
CH_nCl_3	CCl_3	1,1,1-trichloroethane	$1\text{CH}_3, 1\text{CCl}_3$
	CHCl_3	chloroform	1CHCl_3
CCl_4		tetrachloromethane	1CCl_4
ACCl		chlorobenzene	$5\text{ACH}, 1\text{ACCl}$
CH_nNO_2	CHNO_2	2-nitrobutane	$2\text{CH}_3, 1\text{CH}_2, 1\text{CHNO}_2$
	CH_2NO_2	1-nitropropane	$1\text{CH}_3, 1\text{CH}_2, 1\text{CH}_2\text{NO}_2$
	CH_3NO_2	nitromethane	$1\text{CH}_3\text{NO}_2$
ACNO_2		nitrobenzene	$5\text{ACH}, 1\text{ACNO}_2$
CS_2		carbon disulfide	1CS_2
CH_nSH	CH_2SH	ethanethiol	$1\text{CH}_3, 1\text{CH}_2\text{SH}$
	CH_3SH	methanethiol	$1\text{CH}_3\text{SH}$
furfural		furfural	1furfural
$(\text{CH}_2\text{OH})_2$		1,2-ethanediol	$1(\text{CH}_2\text{OH})_2$
I	I	1-iodoethane	$1\text{CH}_3, 1\text{CH}_2, 1\text{I}$
Br	Br	1-bromoethane	$1\text{CH}_3, 1\text{CH}_2, 1\text{Br}$
CH_nC	$\text{C}\equiv\text{C}$	2-pentyne	$2\text{CH}_3, 1\text{CH}_2, 1\text{C}\equiv\text{C}$
	$\text{CH}\equiv\text{C}$	1-butyne	$1\text{CH}_3, 1\text{CH}_2, 1\text{CH}\equiv\text{C}$
	$\text{CH}\equiv\text{CH}$	ethyne	$1\text{CH}\equiv\text{CH}$
DMSO	$(\text{CH}_3)_2\text{SO}$	dimethylsulfoxide	$1(\text{CH}_3)_2\text{SO}$
$\text{CH}_2\equiv\text{CHCN}$		acrylonitrile	$1\text{CH}_2=\text{CHCN}$
$\text{Cl}(\text{C}=\text{C})$	$\text{Cl}(\text{C}=\text{C})$	trichlorethylene	$1\text{CH}=\text{C}, 3\text{Cl}(\text{C}=\text{C})$
ACF	ACF	hexafluorobenzene	6ACF
DMF	DMF-1	dimethylformamide	1DMF-1
	DMF-2	diethylformamide	$2\text{CH}_3, 1\text{DMF-2}$
CF_n	CF	perfluoromethylcyclohexane	$1\text{CH}_3, 5\text{CH}_2, 1\text{CF}$
	CF_2		
	CF_3	perfluorohexane	$2\text{CF}_3, 4\text{CF}_2$
COO	COO	dimethyl oxalate	$2\text{CH}_3, 2\text{COO}$
SiH_2	Si	hexamethyldisiloxane	$6\text{CH}_3, 1\text{SiO}, 1\text{Si}$
	SiH	heptamethyltrisiloxane	$7\text{CH}_3, 2\text{SiO}, 1\text{SiH}$
	SiH_2	diethylsilane	$2\text{CH}_3, 2\text{CH}_2, 1\text{SiH}_2$
	SiH_3	methylsilane	$1\text{CH}_3, 1\text{SiH}_3$
SiO	SiO	octamethylcyclotetrasiloxane	$8\text{CH}_3, 4\text{SiO}$
	SiHO	1,1,3,3-tetramethyldisiloxane	$4\text{CH}_3, 1\text{SiHO}, 1\text{SiH}$
	SiH_2O	1,3-dimethyldisiloxane	$2\text{CH}_3, 1\text{SiH}_2\text{O}, 1\text{SiH}_2$
NMP	NMP	N-methylpyrrolidine	1NMP
CClF	(8 subgroups)	represents freons	
CON	(6 subgroups)	represents amides	
OCCOH	(2 subgroups)	represents glycol ethers	
CH_2S	(3 subgroups)	represents organic sulfides	
morpholine		morpholine	1morpholine
thiophene	(3 subgroups)	represents thiophenes	

^a Ref. 168.
^b A represents an aromatic moiety.

The basic equational form of UNIFAC and many other QSARs is

$$GC = GC_{\text{combinatorial}} + GC_{\text{residual}}$$

where GC is the group contribution, $GC_{\text{combinatorial}}$ is based only on composition and size and shape effects, and GC_{residual} depends on these effects and intermolecular forces. GC s are summed to estimate the activity coefficient or excess Gibbs energy. These GC parts are of logarithmic form consistent with the relationship between excess Gibbs energy and activity coefficients.

UNIQUAC and UNIFAC embody a generalized form of Guggenheim's lattice solution model (28,169) and aspects of Derr and Deal's solution-of-groups theory (32). The former is the basis for the equations governing molecular interactions in UNIFAC; the latter provides the method for assessing group contributions once the groups have been selected. A key assumption in UNIQUAC and UNIFAC is that perturbations in the solution lattice can be equated to excess Gibbs energy and therefore to the activity coefficient by use of appropriate correction factors. Several textbooks on the subject suggest these assumptions are reasonable at ordinary (<500 kPa (5 atm)) pressures and over moderate (300–425 K) temperature ranges for nonelectrolyte substances well below their critical points. A description of the lattice theory of solutions is available (118).

UNIQUAC generalizes Guggenheim's model to mixtures containing molecules of varying sizes by using Wilson's local-composition theory (165) which proposes that when a solution is viewed microscopically, the mixture is not homogeneous with regard to composition. Whereas this assumption may be of little consequence to some bulk fluid properties, it is salient to liquid-mixture modeling. In UNIQUAC, a molecule is divided into equal segments where each segment occupies a lattice intersection, or cell. Guggenheim's ratio of like and unlike neighbors (each neighbor is a molecule occupying one cell) is replaced with local area fractions (ratio of like and unlike segments) to correct for the differences in proximity effects of end vs interior groups. Because all segments in a given molecule are assumed equal (even though chemical differences exist in real molecules), energy parameters for the method represent averages over the individual molecular topologies. Statistical aspects of the lattice structure for UNIQUAC were left the same as Guggenheim's model (169).

Linking this molecular model to observed bulk fluid PVT-composition behavior requires a calculation of the number of possible configurations (microstructures) of a mixture. There is no exact method available to solve this combinatorial problem (28). ASOG assumes the athermal (no heat of mixing) Flory-Huggins equation for this purpose (118,170,171). UNIQUAC claims to have a formula that avoids this assumption, although some aspects of athermal mixing are still present in the model.

UNIQUAC is significant because it provides a means to estimate multicomponent interactions using no more than binary interaction experimental data, bond angles, and bond distances. There is an implicit assumption that the combinatorial portion of the model, ie, the size and shape effects, can be averaged over a molecule and that these can be directly related to molecular surface area and volume. This assumption can be found in many QSAR methods and probably makes a significant contribution to the generally low accuracy of many QSAR prediction techniques.

UNIFAC assumes that the size, shape, and intermolecular interactions can be calculated by stoichiometrically and linearly summing group contributions of these parameters for a given system. Size and shape are related to van der Waals group volume and surface areas (172) and are subsequently summed and averaged over each molecule type. Any steric or relative positional effects are lost in this averaging, but it does allow for the segments to be different from one molecule to the next. This feature is not accommodated by ASOG's Flory-Huggins assumption (170,171). UNIFAC's interaction parameter summation differs from its volume and surface area summations only in that it is formulated as a difference to a reference substance containing only molecules of the same group distribution as the substance from which the interaction parameter data is derived. In this way, activity coefficients become unity for pure substances. Although the theoretical validity of these linear summations may be challenged, this technique does provide a mathematically simple expression.

Most of the assumptions are based on idealized models, indicating the limitations of the mathematical methods employed and the quantity and type of experimental data available. For example, the details of the combinatorial entropy of a binary mixture may be well understood, but modeling requires, in large measure, uniformity so the statistical relationships can be determined. This uniformity is manifested in mixing rules and a minimum number of adjustable parameters so as to avoid problems related to the mathematics, eg, local minima and multiple solutions.

Updates for the UNIFAC method primarily expand the list of groups and group interactions and make small adjustments to some of the experimental constants in the equations (39,173,174). Modifications to mixing rules and the like are not common in the more recent literature, although correction factors and similar features have been proposed for most of the underlying theories. For example, an adjustment to the Flory-Huggins theory correlating molecular surface area-to-volume ratio to combinatorial entropy, making the theory more accurate for nonpolymeric substances has been proposed (175). A degree of mixing correction factor to Wilson's equation has also been provided (176,177), but this correction is not included as a variable parameter in ASOG or UNIFAC, probably in part because of the mathematical and data acquisition difficulties it would create. A strategy for selection of this correction factor is also available (29). Presentations of the Wilson equation (118) often do not show this correction factor, which is commonly called the liquid's coordination number.

Both UNIFAC and ASOG are typically generalized to multicomponent systems in commercial software packages. An important feature of these methods is that only binary interaction information is used to generate multicomponent predictions.

Assessment of UNIFAC. UNIFAC is a method to predict the activity of binary liquid solutions in the absence of all data except structural information. Because state-of-the-art real fluid estimation methods are empirical or semi-empirical, the use of more data results in improved activity estimation.

UNIFAC was developed for situations where experimental data are scarce, and its application should generally be that of last resort (3,7,162,178). That is, UNIFAC is not a method for comparative testing of methods based on experimental data even though the method is sometimes listed in commercial computer simulator menus without indication of its limitations.

Numerous assessments of the reliability of UNIFAC for various applications can be found in the literature. Extrapolating a confidence level for some new problem is ill-advised because accuracy is estimated by comparing UNIFAC predictions to experimental data. In some cases, the data are the same as that used to generate the UNIFAC interaction parameters in the first place. Extrapolating a confidence level for a new problem requires an assumption that the nature of the new problem is similar to that of the UNIFAC test systems previously considered. With no more than structural information, such an assumption may not be valid.

A general review of the UNIFAC and ASOG (32) methods, through the 1985 updates for the American Institute of Chemical Engineers (AIChE) under the Design Institute for Physical Property Data (DIPPR Project 802), is available (179). For temperatures in the range 300–425 K and pressures below 500 kPa (5 atm), UNIFAC is recommended as is the ASOG method (163). Investigation of the availability of interaction parameters is suggested as is comparing errors for the two methods to make a selection for a given problem. A general approach to method selection for specific problems and a comparison of binary and ternary solubility predictions to experimental results for a wide range of functional groups are available (162). In making a method selection, it is important to confirm that the method version matches the selection procedure. This confirmation is not always easy to make, based on commercial software documentation.

UNIFAC's combinatorial term is based on the Stavermann-Guggenheim term (180,181) which is intended to account for size and shape effects of groups. The original ASOG method uses a Flory-Huggins type of approach (170,171). Both methods neglect temperature effects. The ASOG residual term includes temperature dependence, whereas UNIFAC's does not. However, no evidence that this resulted in improved predictions by ASOG has been found (179). Where the residual term is zero, ie, there are no molecular interactions, the Stavermann-Guggenheim term (UNIFAC) gives exaggerated predictions of nonideality (40). No preference was indicated for nonathermal mixtures. The expression for excess Gibbs energy in UNIFAC is not flexible enough to produce reliable predictions for strongly associating or solvating systems in dilute regions (179). Infinite dilution activity coefficients usually exceed 10 for such systems. Predictions for systems containing symmetrical molecules were found to be no better than those by Raoult's law. The use of a simple equation of state having zero interaction coefficients for such situations has been recommended (179). All methods tested, eg, NRTL and UNIQUAC, which use more data, were recommended over UNIFAC for problems where such data were available.

Modified UNIFAC parameters applicable to both vapor-liquid and liquid-liquid environmental problems have been developed (23,24,182,183), extending the UNIFAC range for selected dilute, nonideal systems. Several authors suggest addition of corrector groups to account for steric features, eg, a nonaromatic cyclic group (34). Inclusion of numerous groups containing first members of homologous series (5) suggests the UNIFAC developers were

also aware of this problem.

The use of UNIFAC for estimating activity coefficients in binary and multicomponent organic and organic–water systems is recommended for those systems composed of nonelectrolyte, nonpolymer substances for which only structural information is known. UNIFAC is not recommended for systems for which some reliable experimental data are available. The method, including revisions through 1987 (39), is available in commercial software packages such as AspenPlus (174).

Group Contribution Method Survey. Group contribution methods to estimate physical properties vary in reliability and usage. Almost all are restricted to organic compounds. Before a method is used in studies requiring more than rough estimation, the literature should be reviewed. Concise descriptions of many of these methods including comparisons to nongroup contribution methods are available (5,7,184).

Critical Properties. Several methods have been developed to estimate critical pressure, P_c ; temperature, T_c ; and volume, V_c . Many other properties can be estimated from these properties. Error propagation can be large for physical property estimations based on critical properties from group contribution methods. Thus sensitivity analyses are recommended. The Ambrose method (185) was found to be more accurate (186) than the Lydersen (187) method, although it is computationally more complex. The Joback and Reid method (188) is only slightly less accurate overall than the Ambrose method, and is more accurate for some specific substances. Other methods of lesser overall accuracy are also available (189,190) (T_c , V_c), (191,192) (T_c , P_c), (193) (T_c and (194) (V_c). A second-order method using Benson-type groups (195,196) and normal boiling point for predicting critical temperature and pressure of hydrocarbons has been developed (197). This was extended to include organic compounds containing oxygen, nitrogen, sulfur, and halogens (198).

Parachor is the name (199) of a temperature-independent parameter to be used in calculating physical properties. Parachor is a function of liquid density, vapor density, and surface tension, and can be estimated from structural information. Critical constants for about 100 organic substances have been correlated to a set of equations involving parachors and molar refraction (200).

Normal Boiling Point–Freezing Point. One method's deviation for boiling and freezing point data is reported to be less than 2 K for 80% of a 600 compound set and less than 5 K for an additional 18% (201). However, this method is limited to organic compounds having only one functional group and a short base group such as methyl, vinyl, and phenyl. In contrast, the Joback method (188) has much wider applicability and a reported average error of 12.9 K. The Joback method also includes correlating parameters for freezing point at atmospheric pressure. A 1.29% average absolute error for predictions of a diverse set of 1169 organic species based on a nonlinear group contribution model has been reported (202). The parachor method also estimates normal boiling point (200), and a group contribution based correction factor for critical temperature to estimate boiling point has been developed (187). Reviews of other methods are available (5,19).

Liquid Molar Volume. An average absolute error of approximately 1% for a range of liquid molar volumes of organic solvents and oligomers between their melting and normal boiling points has been reported (203). In addition, a 2.7% average error for selected amorphous polymers was estimated. The Le Bas method (204) estimates liquid molar volume at the normal boiling point. An average error of 3.9% for 32 diverse chemical species has been reported for this method (7). For many species, the Tyn and Calus equation (205), which relates normal boiling point molar volume to V_c , results in more accurate estimations than Le Bas (204) when the Ambrose (185) or Joback (188) method is used to estimate V_c .

Vapor Pressure. Empirical vapor pressure equations have been generalized, based on structure through Kamlet-Taft solvatochromic parameters (206). This correlation falls under the linear solvation energy relationship class of correlations. However, it has been extended to derive vapor pressures from UNIFAC activity coefficient estimations by an imaginary inert methane matrix solvent. An indirect means of vapor pressure prediction has been proposed (207) using, for example, a modified set of UNIFAC groups fit to the Antoine equation constants (208). Halogenated aromatic hydrocarbon vapor pressures have been fit (209) to groups (210,211). Perfluorinated saturated hydrocarbons have been fit (212) to a modified set of Lawson's groups (213), as have 48 substituted benzenes (214).

Heats of Vaporization and Fusion. A simple linear summation of most of the Lydersen groups (187) has been proposed for heat of vaporization at the normal boiling point and heat of fusion at atmospheric pressure for a wide variety of organic compounds (188). Average errors of 1.2 and 4.3% for group contribution-based estimations of heats of vaporization for selected *n*- and iso-alkanes, respectively, have been reported (215).

Heat Capacity. The multiple property estimation methods for constant pressure ideal-gas heat capacities cover a broad range of organic compounds (188,216,217). Joback's method (188) is the easiest to use; however, usage of all these methods has been recommended only over the range 280–1100 K (7). An accurate method for ideal-gas heat capacities (constant pressure), limited to hydrocarbons, has been presented (218) that involves a fit of seven variables, and includes steric, ring, branching, alkene, and even allene corrections.

Constant volume heat capacities for liquid organic compounds were estimated with a four parameter fit (219). A 1.3% average absolute error for 31 selected species was reported. A group contribution method for heat capacities of pure solids and liquids based on elemental composition has also been provided (159).

Surface Tension. The relationship between surface tension and liquid molar volume (220), and the group contribution methods for liquid molar volume can be utilized to estimate surface tension.

Viscosity. A corresponding states method that requires critical pressure, temperature, and dipole moment has been developed for low pressure gas viscosity (221). This method, which includes a group contribution parameter, is also applicable to gas mixtures. Whereas a group contribution method is not available for dipole moment, the influence this parameter has can be neglected for many species.

Pure, low temperature organic liquid viscosities can be estimated by a group contribution method (7) and a method combining aspects of group contribution and connectivity indexes theories (222). Caution is recommended in the use of these methods because the calculated absolute errors are as high as 100% for individual species in a 150-compound, 10-family test set (223). A new method based on a second-order fit of Benson-type groups with numerous steric correctors is suggested as an alternative. Lower errors are claimed for the same test set.

Thermal Conductivity. A method for determining the thermal conductivity of polyatomic organic gases based on critical properties and a series of correction factors based on functional groups has been developed (224). Organic liquid thermal conductivity has also been estimated utilizing a second-order fit of Benson-type groups (225). An average deviation of 5.9% was reported for 266 compounds.

Acentric Factor. A method restricted to alkanes has been proposed for estimating acentric factors (226). A 2.2% average absolute error for 59 species was reported.

Flammability Limits. Some 1358 compounds selected from the DIPPR Compilation File (Pennsylvania State University, 1991; Ref. 4) have been fit for upper and lower flammability limits (227). Average errors reported were 0.266% (volume) and 0.06% (volume) for upper and lower flammability limits, respectively. A detailed analysis by functional group classification is included that identifies classifications with high error for several methods.

Enthalpy of Formation. The methods of several workers (188,216–218) include parameters for estimating enthalpies of formation.

Gibbs (Free) Energy. The methods of several workers (188,218) include parameters for estimating Gibbs energy of formation. In addition, based on its thermodynamic relationship to the activity coefficient, excess Gibbs energy can be estimated through the UNIFAC and ASOG methods.

Entropy. Standard entropies can be estimated by available methods (216–218).

Activity Coefficients. Most activity coefficient property estimation methods are generally applicable only to pure substances. Methods for properties of multicomponent systems are more complex and parameter fits usually rely on less experimental data. The primary group contribution methods of activity coefficient estimation are ASOG and UNIFAC. Of the two, UNIFAC has been fit to more combinations of groups and therefore can be applied to a wider variety of compounds. Both methods are restricted to organic compounds and water.

Henry's Constant. UNIFAC interaction parameters have been refit (227–229) to estimate Henry's law constants for dilute organic-in-water systems. Because this method is based on UNIFAC, it can be used for organic species with multiple function groups. A work following similar lines to the method of Reference 230 was presented for octanol–water partitioning coefficient and water solubility (183). Another UNIFAC-based method applied to methane, nitrogen, and oxygen gas phases has been developed (230). A method not based on UNIFAC applies to hydrocarbons and compounds having single functional groups (231,232). A bond contribution method for a diverse set of 59 organic compounds is also available (233).

Octanol–Water Partition Coefficient. The *f*-Fragment approach (234–236) has been reviewed (227) and another method similar to the UNIFAC refit for Henry’s constant has been proposed. Improved accuracy for many species and the ability to correct for temperature effects have been claimed for the newer method.

Diffusion Coefficient. The method of Reference 237 has been recommended for many low pressure binary gases (238). Other methods use solvent and solute parachors to calculate diffusion coefficients of dissolved organic gases in liquid solvents (239,240). Molar volume and viscosity are also required and may be estimated by the methods previously discussed. Caution should be exercised because errors are multiplicative by these methods.

Second Virial Coefficient. A group contribution method including polar and nonpolar contributions has been proposed for second virial coefficients (241). This method has been applied to both pure components and mixtures, the latter through prediction of cross-second virial coefficients.

Chemical Reactivity. Among other things, reactivity plays a vital role in determining fate of pollutants in the environment. A prototype computer system to estimate reactivity of systems based on structure, SPARC (SPARC Performs Automated Reasoning in Chemistry), is described (242). As of this writing, procedures for hydrolysis rate constants, ionization pK_a uv–vis light absorption spectra, and several physical properties have been developed for SPARC.

Toxicology and Other Biological System Studies. Quantitative structure–relationship (QSAR) studied in the field of toxicology are extensive. These relationships range from functional group correlations to conventional group contributions to solvation energy relationships to molecular connectivity indexes. The U.S. Fish and Wildlife Service has developed an expert system that generates linear solvation energy relationship (LSER) parameters based on contaminant chemical structure and uses them for predicting aquatic toxicity (243). Overviews of and specific QSAR uses in environmental toxicology were presented in 1983 (244–247). An overview of QSAR studies related to drugs acting on the central system has also been compiled (248). A historical review of research into the relationship of biological activity and aqueous solubility to partition coefficient as well as discussions of related QSAR-based models is available (249) as is a strategy for ranking chemical hazards based on QSAR techniques (250). Specific property studies on acute fish toxicity relationships can be found in the literature (251,252). Other studies propose QSARs for specific chemical classes, such as the organochlorine and synthetic pyrethroid insecticide study (253).

Linear Free Energy–Linear Solvation Energy Relationships. Linear free energy (LFER) and linear solvation energy (LSER) relationships are used to develop correlations between selected properties of similar compounds. These are fundamentally a collection of techniques whereby properties can be predicted from other properties for which linear dependency has been observed. Linear relationships include not only simple $y = mx + b$ relationships, but also more complicated expressions such as the Hammett equation (254) which correlates equilibrium constants for benzenes,

$$\log(K_x/K_H) = \rho\sigma_x$$

where K_x and K_H are equilibrium constants for substituted and unsubstituted benzenes, respectively, ρ is a reaction parameter, and σ_x is a substituent parameter.

The literature is not particularly clear on the distinction between LFERs and LSERs. In some instances, the terms are used interchangeably. For the present, consider LFERs as those relationships for which properties are linearly related to free energy. These include equilibrium constants, reaction rates, various spectral properties, polarographic half-wave lengths, dipole moments, and redox behavior (255). These have been used extensively to investigate the relationship between structure and chemical, physical, and biological properties of organic compounds, in particular. LSERs can be thought of as that subset of LFERs concerned with properties relating to solubility and partitioning. An active area of LSER research involves relating solubility-related parameters, eg, 1-octanol–water partitioning coefficient, K_{ow} , to the evaluation of the hazards of environmental contaminants (see ENVIRONMENTAL IMPACT).

The distinction between LFERs and group contribution methods is also confusing. Many group contribution methods can be considered to be simply extensions of LFERs for which the independent parameters are structural. From this perspective, both group contribution techniques and LSERs are subsets of LFERs. Another viewpoint holds that the differences are largely semantical, owing to the historical fact that many group contribution techniques originate in the works of thermodynamicists and engineers, whereas many LFER–LSER methods are first found in the writings of chemists, biologists, and environmental scientists.

An excellent historical review of LFERs is available (255). An example of an LFER study, relating partition coefficients to aqueous solubility of organic liquids, may be found in Reference 256.

Kamlet-Taft Linear Solvation Energy Relationships. Most recent works on LSERs are based on a powerful predictive model, known as the Kamlet-Taft model (257), which has provided a framework for numerous studies into specific molecular thermodynamic properties of solvent–solute systems. This model is based on an equation having three conceptually explicit terms (258).

$$Prop = \text{cavity term} + \text{dipolar term} + \text{hydrogen-bonding terms} \tag{105}$$

where *Prop* is the physical property of interest, the *cavity term* describes the energy associated with creation and stabilization of a hole of correct size and shape in a continuous solvent phase large enough to accept a solute molecule, and the remaining terms relate electrostatic and hydrogen-bonding interactions between the molecules to stabilize the solute molecule in the hole. Such an equation is readily transformed into a form amenable to a quantitative structure–activity approach such as,

$$Prop = Prop_0 + mV_i/100 + s\pi^* + b\beta_m + a\alpha_m \tag{106}$$

where *m*, *s*, *b*, and *a* are constants for a particular property and set of conditions. The normally endoergic cavity term, $mV_i/100$, contains $V_i/100$, the van der Waals molecular volume which is scaled to units consistent with the other terms. The dipolar term, $s\pi^*$, is a measure of the largely exoergic solute–solvent dipole–dipole and dipole-induced dipole effects interactions, and π^* is a measure of the ability of nonspecific dielectric effects to stabilize neighboring charges and dipoles. The remaining terms incorporate the exoergic effects of the formation of hydrogen bonds. The B_m parameter is used for conditions in which the solvent is a hydrogen bond donor acid (HBD) and the solute is a hydrogen bond acceptor base (HBA), whereas α_m involves an HBA solvent and an HBD solute. The subscript *m* indicates that only the monomeric solute is considered, irrespective of the fact that it may self-associate. These last terms form the framework for expanded definitions of acids and bases which include nonelectrolyte systems. This approach has been used to examine solvent effects on equilibria and reaction rates, the influence of solute effects by water, blood, etc, solubilities, phase distribution, and carbon adsorption (259).

Owing to the original determination from uv–vis spectral solvatochromic shifts, π^* , B_m , and α_m are called solvatochromic parameters. General rules for estimation of these variables have been proposed (258). Examples of individual parameter investigations are available (260,261). As previously mentioned, individual LFER–LSER studies are performed on related materials. A common method to link these individual studies to group contribution methods, and thereby expand the applicability, is by expansion of solvatochromic parameters to log–linear relationships, such as

$$\pi_x = \log_{10} P_x - \log_{10} P_H$$

where P_x is a coefficient related to a derivative compound, P_H is related to its parent molecule, and π_x is the resultant solvatochromic parameter or solvation property. As defined (262), the subscript *x* refers to the π value for a substitute group. This so-called parent–derivative method has been used for estimation of octanol–water partition coefficients for sets of similar compounds.

A sampling of applications of Kamlet-Taft LSERs include the following. (1) The Solvatochromic Parameters for Activity Coefficient Estimation (SPACE) method for infinite dilution activity coefficients where improved predictions over UNIFAC for a database of 1879 critically evaluated experimental data points has been claimed (263). (2) Observation of inverse linear relationship between log 1-octanol–water partition coefficient and liquid

(or subcooled liquid) solubility (264). (3) Correlation of fish bioconcentration factors with solvatochromic parameters (265). (4) The distribution of organic solutes between water and immiscible organic solvents (266). (5) Carbon adsorption of organic compounds (267). A similar, alternative approach is one based on solvophobic theory (244). (7) Correlation of octanol–water partition coefficients of organic nonelectrolytes (including strong HBD solutes) (268).

Molecular Connectivity Indexes and Graph Theory. Perhaps the chief obstacle to developing a general theory for quantification of physical properties is not so much in the understanding of the underlying physical laws, but rather the inability to solve the requisite equations. The plethora of assumptions and simplifications in the statistical mechanics and group contribution sections of this article provide examples of this. Computational procedures are simplified when the number of parameters used to describe the salient features of a problem is reduced. Because many properties of molecules correlate well with structures, parameters have been developed which grossly quantify molecular structural characteristics. These parameters, or connectivity indexes, are usually based on the numbers and orientations of atoms and bonds in the molecule.

A forerunner to recent connectivity indexes is found in work on the properties of saturated hydrocarbons (269,270). The Wiener index, ω , can be calculated by taking the product of the number of carbon atoms on the two sides of each C—C bond and then summing these products for all C—C bonds in the molecule (271). For example, 2-methylpentane, 2,2-dimethyl butane, and 2,3-dimethyl-butane have Wiener indexes of 32, 28, and 26, respectively. It can be readily seen that ω combines both molecular bulk and branching in a single parameter. As is to be expected, a single parameter does not generally provide sufficient information to do accurate property estimation. By adding a second index, p , which is equal to the number of pairs of carbon atoms separated by three bonds and normalizing ω by the square of the number of carbon atoms in the molecule (N^2), a general formula for estimating several properties was generated (269,270):

$$P_i = a(\omega / N^2) + bp + c \tag{107}$$

where a , b , and c are constants appropriate to property P_i .

A series of connectivity indexes based on this same type of information, but having mathematical form more closely aligned with potential functions, has also been developed (272–274). This approach allows for a series of indexes to be calculated based on assemblages of increasing numbers of contiguous bonds and has been successfully applied to a variety of biological and physical properties (275), including: boiling point (274,276), density (274), heats of formation (273), molar refractive index (275), partition coefficients (277) and polarity of solvents (278), aqueous solubility (45), surface tension (279), and critical volume and acentric factor (280).

As computing capability has improved, the need for automated methods of determining connectivity indexes, as well as group compositions and other structural parameters, for existing databases of chemical species has increased in importance. New naming techniques, such as SMILES, have been proposed which can be easily translated to these indexes and parameters by computer algorithms. Discussions of the more recent work in this area are available (281,282). SMILES has been used to input Contaminant structures into an expert system for aquatic toxicity prediction by generating LSER parameter values (243,258).

Perhaps a more flexible approach to incorporating structural information into a series of rule-based parameters is found in the incorporation of graph theory into chemical representation studies. Graph theory, introduced in 1736, is based on ideas from topology and combinatorics (271). Introductions to this branch of mathematics are available (283,284). The enhanced flexibility of graph theory stems from the fact that it retains molecular structural information in a matrix format which is essentially a two- (or more) dimensional representation of a molecule's spatial orientation. Through the use of large computer capacity and sophisticated regression techniques, parameters of optimum correlation to the property of interest may be obtained. An excellent introduction to the application of graph theory to structure–property relationships is available (285). An example of pertinent calculations related to alkanes may be found in Reference 286.

Whereas graph theory and other computer-intensive approaches can provide insights into the area of physical property correlation previously unobtainable, the general guideline should be to use no more parameters than are necessary to achieve the desired level of accuracy in the results. Many so-called ad hoc descriptors, which are usually based on few fitting parameters, have been successfully used (287–291). For example, it has been suggested (271) that two indexes, the number of carbon atoms and the number of terminal methyl groups, can be combined with three fitting parameters in a linear equation for alkane normal boiling points. A correlation coefficient of 0.987 for 39 species has been reported. LSER (243,258) uses only four variables, and a routine correlation coefficient range of 0.93–0.98 has been reported for numerous investigations. Concise introductions to the basic concepts of predicting chemical and physical properties from molecular topology may be found in the literature (271,292).

BIBLIOGRAPHY

"Data Correlation" in *ECT* 1st ed., Vol. 4, pp. 846–873, by D. F. Othmer and E. S. Roszkowski, Polytechnic Institute of Brooklyn; "Data—Interpretation and Correlation Correlation" in *ECT* 2nd ed., Vol. 6, pp. 705–755, by D. F. Othmer, Polytechnic Institute of Brooklyn; "Engineering and Chemical Data Correlation" in *ECT* 3rd ed., Vol. 9, pp. 45–117, by D. F. Othmer, Polytechnic Institute of New York.

1. D. M. Himmelblau, *Basic Principles and Calculations in Chemical Engineering*, 5th ed., Prentice Hall, Inc., Englewood Cliffs, N.J., 1989.
2. M. Modell and R. C. Reid, *Thermodynamics and Its Applications*, 5th ed., Prentice Hall, Inc., Englewood Cliffs, N.J., 1974.
3. J. M. Smith and H. C. Van Ness, *Introduction to Chemical Engineering Thermodynamics*, 4th ed., McGraw-Hill Book Co., Inc., New York, 1987.
4. *Manual for Predicting Chemical Process Design Data*, Design Institute for Physical Property Data (DIPPR), AIChE, 1983 (updated in 1986).
5. W. J. Lyman, W. F. Reehl, and D. H. Rosenblatt, *Handbook of Chemical Property Estimation Methods*, McGraw-Hill Book Co., Inc., New York, 1982.
6. *EPA Superfund Public Health Evaluation Manual*, EPS 540 1-86 060, U.S. Environmental Protection Agency, Office of Emergency and Remedial Response, Washington, D.C., October 1986.
7. R. C. Reid, J. M. Prausnitz, and B. E. Poling, *The Properties of Gases and Liquids*, 4th ed., McGraw-Hill Book Co., Inc., New York, 1987.
8. R. H. Perry and C. H. Chilton, *Chemical Engineer's Handbook*, 5th Ed., McGraw-Hill Book Co., Inc., New York, 1973.
9. D. R. Lide, ed., *CRC Handbook of Chemistry and Physics*, 71st ed. CRC Press, Boca Raton, 1990.
10. S. M. Walas, *Phase Equilibria in Chemical Engineering*, Butterworth Publishers, Boston, 1985.
11. L. Berenyi, *Z. Phys. Chem.* **94**, 628 (1920).
12. L. Berenyi, *Z. Phys. Chem.* **105**, 55 (1923).
13. M. M. Dubinin, *Chem. Rev.* **60**, 235–241 (1960).
14. M. M. Dubinin and L. V. Radushkevich, *Dokl. Akad. Nauk. SSSR* **55**, 327, (1947).
15. M. Polanyi, *Verh. Dtsch. Phys. Ges.* **16**, 1012 (1914).
16. M. Polanyi, *Z. Elektrochem.* **26**, 371 (1920).
17. A. L. Myers and S. Sircar, in M. J. McGuire and I. H. Suffet, eds., *Treatment of Water by Granular Activated Carbon*, American Chemical Society, Washington, D.C., 1983, pp. 63–76.
18. J. M. Prausnitz, *Molecular Thermodynamics of Fluid-Phase Equilibria*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1969.
19. R. C. Reid and T. K. Sherwood, *The Properties of Gases and Liquids: Their Estimation and Correlation*, 2nd ed., McGraw-Hill Book Co., Inc., New York, 1966.
20. K. S. Pitzer and R. F. Curl, *J. Amer. Chem. Soc.*, 79 (May 20, 1957).
21. P. W. Atkins, *Physical Chemistry*, W. H. Freeman and Co., San Francisco, 1978, pp. 204–208.
22. C. L. Yaws, *Thermodynamic and Physical Property Data*, Gulf Publishing Co., Houston, Tex., 1992.
23. W. B. Arbuckle, *Env Sci Tech* **17**(9), 537–542 (1983).
24. W. B. Arbuckle, *Env Sci Tech.* **20**(10), 1060–1064 (1986).

25. G. M. Wilson and C. H. Deal, *Ind. Eng. Chem. Fund.* **1**(1), 20–23 (1962).
26. J. H. Hildebrand, J. M. Prausnitz, and R. L. Scott, *Regular and Related Solutions*, Van Nostrand Reinhold, Co., Inc., New York, 1970.
27. F. A. Mato, R. B. Mato, and F. Mato, *Ind. Eng. Chem. Res.* **28**(9), 1441–1446 (1989).
28. D. S. Abrams and J. M. Prausnitz, *AIChE J.* **21**, 116 (1975).
29. M. Hiranuma, *J. Chem. Eng. Japan* **8**(2), 162–163 (1975).
30. I. Nagata, M. Nagashima, and M. Ogura, *J. Chem. Eng. Japan* **8**(5), 406–408 (1975).
31. T. Tsuboka and T. Katayama, *J. Chem. Eng. Japan* **8**(3), 181–187 (1975).
32. E. L. Derr and C. H. Deal, *I. Chem. E. Symposium Series* **32**, 3:40–3:51 (1969).
33. J. H. Park and P. W. Carr, *Anal. Chem.* **59**(21), 2596–2602 (1987).
34. E. R. Thomas and C. A. Eckert, *Ind. Eng. Chem. Process Des. Dev.* **23**(2), 194–209 (1984).
35. Aa. Fredenslund, R. L. Jones and J. M. Prausnitz, *AIChE J.* **21**, 1086–1099 (1975).
36. S. Skjold-Jorgensen, B. Kolbe, J. Gmehling, and Rasmussen, *Ind. Eng. Chem. Process Des. Dev.* **18**(4) 714–722 (1979).
37. J. Gmehling, P. Rasmussen, and Aa. Fredenslund, *Ind. Eng. Chem., Process Des. Dev.* **21**(1), 118–127 (1982).
38. E. A. Macedo, U. Weidlich, J. Gmehling, and P. Rasmussen, *Ind. Eng. Chem. Process Des. Dev.* **22**(4), 676–678 (1983).
39. D. Tiegs and co-workers, *Ind. Eng. Chem. Res.* **26**(1), 159–161 (1987).
40. I. Kikic, P. Alessi, P. Rasmussen, and Aa. Fredenslund, *Can. J. Chem. Eng.* **58**, 253–258 (Apr. 1980).
41. O. Sinanoglu in B. Pullman, ed., *Molecular Associations in Biology*, Academic Press, New York, 1968, p. 427.
42. N. N. Nirmalakhandan and R. E. Speece, *Env. Sci. Tech.* **22**(11), 1349–1357 (1988).
43. G. Belfort, G. L. Altshuler, K. K. Thallam, C. P. Feerick, Jr., and K. L. Woodfield, *AIChE J.*, **30**(2), 197 (Mar. 1984).
44. N. N. Nirmalakhandan and R. E. Speece, *Env. Sci. Tech.* **24**(4), 575 (1990).
45. N. N. Nirmalakhandan and R. E. Speece, *Env. Sci. Tech.*, **22**(3), 328–338 (1988).
46. N. N. Nirmalakhandan and R. E. Speece, *Env. Sci. Tech.* **3**, 708–713 (1989).
47. J. J. Carroll, *Chem. Eng. Prog.* (Sept. 1991).
48. R. A. Ashworth, G. B. Howe, M. E. Mullins, and T. N. Rogers, *J. Hazard. Matls.* **18**, 25–36 (1988).
49. G. J. Pierotti, C. H. Deal, and E. L. Derr, *Ind. Eng. Chem.* **51**(1), 95–102 (1959).
50. J. M. Gossett and A. H. Lincoff, *Solute Gas Equilibria in Multi-organic Aqueous Systems*, report 1981, AFOSR-TR-81-0858; order no. AD-A109082, NTIS, 1981.
51. J. M. Gossett, *Env. Sci. Tech.* **21**(2), 202–208 (1987).
52. C. Black, G. G. Joris, and H. S. Taylor, *J. Chem. Phys.* **16**(5), 537–543 (1948).
53. R. Stephenson, J. Stuart, and M. Tabak, *J. Chem. Eng. Data* **29**, 287–290 (1984).
54. R. Stephenson and J. Stuart, *J. Chem. Eng. Data* **31**, 56–70 (1986).
55. K. Kinoshita and H. Ishikawa, *Bull. Chem. Soc. Japan* **31**(9), 1081–1082 (1958).
56. W. Y. Shiu and D. Mackay, *J. Phys. Chem. Ref. Data* **15**(2), 911–929.
57. S. Banerjee, S. H. Yalkowsky, and S. C. Valvani, *Env. Sci. Tech.* **14**(10), 1227–1229 (1980).
58. M. M. Miller, S. Ghodbane, S. P. Wasik, Y. B. Tewari, and D. E. Martire, *J. Chem. Eng. Data* **29**, 184 (1984).
59. V. N. Kabadi and R. P. Danner, *Hydrocarbon Proc.* **58**(5), 245–246 (1979).
60. C. L. Yaws and X. Pan, *Oil & Gas J.* 79–82, (Apr. 8, 1991).
61. B. I. Lee and M. G. Kesler, *AIChE J.* **21**, 3 (May 1975).
62. J. R. Brock and R. B. Bird, *AIChE J.* **1**, 2 (June 1955).
63. B. F. Dodge, *Inc. Eng. Chem.* **24**, 12 (Dec. 1932).
64. W. K. Lewis, *Ind. Eng. Chem.*, **28**, 257–262 (Feb. 1936).
65. W. K. Lewis and C. D. Luke, *Trans. Am. Soc. Mech. Eng.* PME-**54**, 8 (1932).
66. W. K. Lewis and C. D. Luke, *Ind. Eng. Chem.* **25**, 7 (July 1933).
67. F. D. Maslan and T. M. Littman, *Ind. Eng. Chem.* **45**, 7 (July 1953).
68. H. P. Meissner and R. Seferian, *Chem. Eng. Prog.* **47**, 11 (Nov. 1951).
69. L. C. Nelson and E. F. Obert, *Trans. Am. Soc. Mech. Eng.* 76 (Oct. 1954).
70. L. Nelson and E. F. Obert, *AIChE J.* **1**, 1 (Mar. 1955).
71. L. Obert, *Ind. Eng. Chem.* **40**, 11 (Nov. 1948).
72. G. J. Su and C. H. Chang, *Ind. Eng. Chem.* **38** (1946).
73. Meslin, *Comptes Rendus cxxvi* (1893).
74. K. S. Pitzer, *J. Chem. Phys.* **7** (Aug. 1939).
75. K. S. Pitzer, *J. Amer. Chem. Soc.* **77**, 13 (July 16, 1955).
76. T. W. Leland, Jr. and P. S. Chapplear, *Ind. Eng. Chem.* **60**, 7 (July 1968).
77. O. Redlich and J. N. Kwong, *Chem. Rev.* **44**, 233–244 (1949).
78. H. G. Rackett, *J. Chem. Eng. Data* **15**, 4 (1970).
79. K. S. Pitzer, *J. Amer. Chem. Soc.* **77**, 13 (July 16, 1955).
80. R. L. Halm and L. Stiel, *AIChE J.*, **3**, 2 (Mar. 1967).
81. G. K. Stipp, S. D. Bai, and L. I. Stiel, *AIChE J.*, **19**, 6 (Nov. 1973).
82. B. I. Lee and M. G. Kesler, *AIChE J.* **21**, 3 (May 1975).
83. R. H. Newton, *Ind. Eng. Chem.* **27**, 3 (Mar. 1935).
84. C. W. Selheimer and co-workers, *Ind. Eng. Chem.* **24**, 5 (May 1932).
85. L. Raidel, *Chemie Ingenieur Technik*, (in German), VCH, Weinheim, 1956, p. 28.
86. R. Danner and T. Daubert, *DIPPR 802 Data Prediction Manual*, AIChE. University Park, Pa., 1986, pp. 7A1–7B2.
87. J. R. Brock, and R. B. Bird, *AIChE J.* **1**, 174 (1955).
88. D. I. Hakim, D. Steinberg, and L. I. Stiel, *Ind. Eng. Chem. Fundam.* **10**(1), 174–175 (1971).
89. M. D. Croucher and M. L. Hair, *J. Phys. Chem.* **81**, 17 (1977).
90. K. Lucas, *Phase Equilibrium and Fluid Properties in the Chemical Industry*, Dechema, Frankfurt, 1980.
91. L. I. Stiel and G. Thodos, *Progress in International Research on Thermodynamics and Transport Properties*, American Society of Mechanical Engineers, Academic Press, Inc., New York, 1962.
92. C. Tsonopoulos, *AIChE J.* **20**, 2 (Mar. 1974).
93. S. Chidambaram and G. Narismhan, *Chem. Eng.* 135, (Nov. 23, 1964).
94. R. D. Gunn, T. Yamada, and D. Whitman, *AIChE J.* **20**(5), 906 (Sept. 1974).
95. H. C. Chen and S. H. Chen, *Ind. Eng. Chem. Fund.* **24**(2), 183 (1985).
96. W. C. Rheinbolt, *Numerical analysis of Parameterized Nonlinear Equations*, Wiley-Interscience, New York, 1986.
97. J. H. Wegenstien, *Commun. ACM.* **1** (1958).
98. O. Orbach and C. M. Crowe, *Can. J. Chem. Eng.* **68**(7–8), 1033–1042 (1971).

99. D. Marquard, *SLAM J. App. Math.*, **11** (1963).
100. J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Non-Linear Equations*, Prentice-Hall, Englewood Cliffs, N.J., 1983.
101. M. J. D. Powell, *Numerical Methods for Non-Linear Algebraic Equations*, Gordon and Breach, New York, 1970.
102. W. J. Lin, J. D. Seader, and T. L. Wayburn, *AIChE. J.* **33**(6), 886–897 (1987).
103. D. S. Davis, *Empirical Equations and Nomography*, McGraw-Hill Book Co., Inc., New York, 1943, pp. 91–160.
104. M. Kraitchik, *Alignment Charts Construction and Use*, D. Van Nostrand Co., Inc., New York, 1944.
105. J. Lipka, *Graphical and Mechanical Computation, Part I. Alignment Charts*, John Wiley & Sons, Inc., New York, 1921.
106. W. C. Marshall, *Graphical Methods for Schools, Colleges, Statisticians, Engineers and Executives*, McGraw-Hill Book Co., Inc., New York, 1921, pp. 167–185.
107. J. B. Peddle, *The Construction of Graphical Charts*, McGraw-Hill, London, 1910.
108. A. S. Levins, *Nomography*, 2nd ed., Fearon Publishers, Belmont, Calif., 1971.
109. J. Angel, *Chem. Eng. Progress* **86**, 85–87 (1990).
110. M. E. Davis, *Numerical Methods for Chemical Engineers*, John Wiley & Sons, Inc., New York, 1984.
111. E. Aarts, and J. Korst, *Simulated Annealing and Boltzmann Machines*, John Wiley & Sons, Inc., New York, 1989.
112. J. A. Nelder and R. Mead, *Comput. J.* **7**, 308 (1965).
113. J. A. Solodar, *Chemtech.* **16**, 372–377 (1986).
114. P. K. Mookerjee, *Chemtech.* **15**, 606–610 (1985).
115. R. J. Amarger, L. T. Biegler, and I. E. Grossman, *Computers Chem. Eng.* **16**, 623–636 (1992).
116. J. J. Ferrada and J. M. Holmes, *Chem. Eng. Progress* **86**, 34–41 (1990).
117. J. A. Fay, *Molecular Thermodynamics*, Addison-Wesley, Reading, Mass. 1965, Chapt. 1, p. 83.
118. J. M. Prausnitz, R. N. Lichtenthaler, and E. G. de Azevedo, *Molecular Thermodynamics of Fluid-Phase Equilibria*, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1986, Chaps. 4, 5, 7; App. III, pp. xv, 6–7, 306–327, 360–364, 509–520.
119. L. Boltzmann, *Lectures on Gas Theory*, English translation University of California Press, 1964); German original published by J. A. Barth, Leipzig, 1896, Part I; 1898, Part II.
120. L. Loeb, *Kinetic Theory of Gases*, 2nd ed., Dover, 1934.
121. J. O. Hirschfelder, C. F. Curtis, and R. B. Bird, *Molecular Theory of Gases and Liquids*, John Wiley & Sons, Inc., New York, 1954.
122. F. W. Sears and G. L. Salinger, *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*, 3rd ed., Addison-Wesley, Reading, Mass. 1975, Chaps. 9. and 11; pp. 320–322.
123. I. Møller, *Thermodynamics*, Pittman Advanced Publishing Program, Boston, 1985, Chapt. 5.
124. J. D. Lewins, ed., *Teaching Thermodynamics*, Plenum Press, New York, 1985.
125. J. W. Whalen, *Molecular Thermodynamics: A Statistical Approach*, John Wiley & Sons, Inc., New York, 1991.
126. E. Schrödinger, *Statistical Thermodynamics*, 2nd ed., Cambridge University Press, Cambridge, UK, 1964.
127. F. Daniels and R. A. Alberty, *Physical Chemistry*, 4th ed., John Wiley & Sons, Inc., New York, 1975, pp. 396–422.
128. W. Pauli, Jr., *Z. Physik* **31**, 765–783 (1925).
129. A. Einstein, *Akademie der Wissenschaften*, Sitzungsberichte, Berlin, 1924, p. 261; 1925, p. 3.
130. E. Fermi, *Z. Physik* **36**, 902 (1926).
131. P. A. M. Dirac, *Proceedings of the Royal Society of London*, A112, 661 (1926).
132. C. Kittel, *Thermal Physics*, John Wiley & Sons, Inc., New York, 1969, Chapt. 9.
133. D. A. McQuarrie, *Statistical Thermodynamics*, Harper & Row, New York, 1973, pp. 164–169.
134. W. Heisenberg, *Z. Physik* **33**, 879–893 (1925).
135. M. Born and P. Jordan, *Z. Physik* **34**, 858–858 (1925).
136. W. Pauli, Jr., *Z. Physik* **36**, 336–363 (1926).
137. E. L. Knuth, *Introduction to Statistical Thermodynamics*, McGraw-Hill, New York, 1966, Chapt. 2.
138. J. Kestin, and J. R. Dorfman, *A Course in Statistical Thermodynamics*, Academic Press, New York, 1971, Chaps. 7–11, 14.
139. F. C. Andrews, *Equilibrium Statistical Mechanics*, John Wiley & Sons, Inc., New York, 1975.
140. N. O. Smith, *Elementary Statistical Thermodynamics—A Problems Approach*, Plenum Press, New York, 1982.
141. R. J. Finkelstein, *Thermodynamics and Statistical Physics—A Short Introduction*, W. H. Freeman, San Francisco, 1969.
142. B. H. Lavenda, *Nonequilibrium Statistical Thermodynamics*, John Wiley & Sons, Inc., Chichester, England, 1985.
143. H. J. Kruezer, *Nonequilibrium Thermodynamics and its Statistical Foundations*, Clarendon Press, Oxford, 1981.
144. H. Ziegler, *An Introduction to Thermomechanics*, 2nd ed. North-Holland Publishing, Amsterdam, distributed in the United States by Elsevier Science Publishing, New York, 1983.
145. E. A. Mason and T. H. Spurling, in *The International Encyclopedia of Physical Chemistry and Chemical Physics*, Vol. 2, topic 10, Pergammon Press, New York, 1969.
146. G. Herzberg, *Atomic Spectra and Atomic Structure*, Prentice Hall, Englewood Cliffs, N.J., 1937, reprinted by Dover, New York, 1944.
147. G. Herzberg, *Molecular Spectra and Molecular Structure, I. Spectra of Diatomic Molecules*, 2nd ed., Van Nostrand, Princeton, N.J., 1950.
148. G. Herzberg, *Infrared and Raman Spectra of Polyatomic Molecules*, Van Nostrand, Princeton, N.J., 1945.
149. E. M. Sparrow and R. D. Cess, *Radiation Heat Transfer*, Brooks Cole Publishing, Belmont, Calif., 1966.
150. D. Q. Kern, *Process Heat Transfer*, McGraw-Hill, New York, 1950.
151. C. Kittel, *Solid State Physics*, 3rd ed., John Wiley & Sons, Inc., New York, 1967, Chapt. 14.
152. G. H. Wannier, *Statistical Physics*, John Wiley & Sons, Inc., New York, 1966, Chapt. 15.
153. A. Einstein, *Investigations of the Theory of the Brownian Movement*, Dover, New York, 1956.
154. P. C. Hiemenz, *Principles of Colloid and Surface Chemistry*, Marcel Dekker, New York, 1977, pp. 106–115.
155. F. Penotti, J. Gerratt, D. L. Cooper, and M. Raimondi, *THEOCHEM* **46**, 421–436 (1988).
156. D. E. Williams, *J. Comput. Chem.* **9**, 745–763 (1988).
157. J. Niwa, *Bull. Chem. Soc. Jpn.* **62**, 226–233 (1989).
158. N. Koga and K. Morokuma, *J. Phys. Chem.* **94**, 5454–5462 (1990).
159. J. E. Hurst, Jr. and B. K. Harrison, *Chem. Eng. Comm.* **112**, 21 (1992).
160. R. M. Felder and R. W. Rousseau, *Elementary Principles of Chemical Processes*, 2nd ed., John Wiley & Sons, Inc., New York, 1986, p. 352.
161. D. M. Himmelblau, *Basic Principles and Calculations in Chemical Engineering*, 3rd ed., Prentice-Hall, Englewood Cliffs, N.J., 1974, p. 270.
162. R. P. Danner, and T. E. Daubert, *Manual for Predicting Chemical Process Design Data*, Design Institute for Physical Property Data (AIChE), 1986, Chapt. 6.
163. K. Kojima and K. Tochigi, *Prediction of Vapor-Liquid Equilibria by the ASOG Method*, Elsevier and Kodansha Ltd., New York, 1979.
164. R. R. Dreisbach, *Pressure–Volume–Temperature Relationships of Organic Compounds*, Handbook Publishers, Sandusky, Ohio, 1952, pp. 294.
165. G. M. Wilson, *J. Am. Chem. Soc.* **86**, 127 (1964).
166. H. Renon and J. M. Prausnitz, *AIChE. J.* **14**, 135 (1968).
167. A. Fredenslund, J. Gmehling, and P. Rasmussen, *Vapor–Liquid Using UNIFAC, a Group Contribution Method*, Elsevier Scientific Publishing,

- Amsterdam, 1977.
168. H. K. Hansen, P. Rasmussen, and A. Fredenslund, *Ind. Eng. Chem. Res.* **30**, 2355–2358 (1991).
 169. E. A. Guggenheim, *Mixtures*, Oxford University Press, 1952.
 170. M. L. Huggins, *J. Phys. Chem.* **9**, 440 (1941); *Ann. N.Y. Acad. Sci.* **43**, 1 (1942).
 171. P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, N.Y., 1953.
 172. A. Bondi, *Physical Properties of Molecular Liquids, Crystals and Glasses*, John Wiley & Sons, Inc., New York, 1968.
 173. T. Magnussen, P. Rasmussen, and A. Fredenslund, *Ind. Eng. Process Dev.* **20**, 331–339 (1980).
 174. Aspen Technology, Inc., *AspenPlus Users Guide*, Cambridge, Mass., 1988, pp. 5–46, appendix pp. C25–C34.
 175. R. N. Lichtenthaler, D. S. Abrams and J. M. Prausnitz, *Can. J. Chem.* **51**, 3071 (1973).
 176. G. Scatchard and G. M. Wilson, *J. Am. Chem. Soc.* **86**, 133 (1964).
 177. H. Renon and J. M. Prausnitz, *AIChE. J.* **15**, 785 (1969).
 178. R. F. Wilcox, *Chem. Eng.* 66 (Oct 17, 1983).
 179. P. A. Gupte, M. Nagvekar, R. P. Danner, and T. E. Daubert, *Documentation of the Basis for Selection of the Contents of Chapter 6 Phase Equilibrium in Manual for Predicting Chemical Process Design Data*, Design Institute for Physical Property Data (AIChE), (1987).
 180. E. A. Guggenheim, *Applications of Statistical Mechanics*, Oxford University Press, London, 1966.
 181. A. J. Staverman, *Recl. Trav. Chim. Pays-Bas* **69**, 163 (1950).
 182. T. N. Rogers and G. Howe, *A UNIFAC Model for Predicting Henry's Law Constants*, Research Triangle Institute, unpublished report, 1988.
 183. G. Wienke and J. Gmehling, *Prediction of Properties of Environmental Interest Using UNIFAC, Documentation from AIChE Annual Meeting presentation*, Miami Beach, Fla., Nov. 1–6, 1992 (available from author at Universitat Oldenburg, Technische Chemie, Postfach 25 03, D-2900 Oldenburg, Germany).
 184. C. Jochum, M. G. Hicks, and J. Sunkel, eds., *Physical Property Prediction in Organic Chemistry, Proceedings of the Beilstein Workshop, May 16–20, 1988, Schloss Korb, Italy*, Springer-Verlag, Berlin and New York, 1988.
 185. D. Ambrose, *Correlation and Estimation of Vapor-Liquid Critical Properties. I. Critical Temperatures of Organic Compounds*, National Physical Laboratory, Teddington, UK, NPL Report 92 (1978, corrected 1980).
 186. K. M. Klinecicz, and R. C. Reid, *AIChE J.* **30**, 137 (1984).
 187. A. L. Lydersen, *Estimation of Critical Properties of Organic Compounds*, University of Wisconsin, Eng. Exp. Sta. Report 3, Madison (1955).
 188. K. G. Joback, and R. C. Reid, *Chem. Eng. Comm.* **57**, 233 (1987).
 189. R. F. Fedors, *AIChE. J.* **25**, 202 (1979).
 190. R. F. Fedors, *Chem. Eng. Comm.* **16**, 149 (1982).
 191. L. Reidel, *Z. Electrochem.* **53**, 222 (1949).
 192. L. Reidel, *Chem. Ing. Tech.* **24**, 353 (1952).
 193. W. R. Gambill, *Chem. Eng.*, **66**(12), 181 (1959); **66**(14), 157 (1959).
 194. A. Vetere, *AIChE J.* **22**, 950 (1976); Errata, **23**, 406 (1977).
 195. S. W. Benson, and J. H. Buss, *J. Chem. Phys.* **29**, 546 (1958).
 196. S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw, and R. Walsh, *Chem. Rev.* **69**(3), 279–324 (1969).
 197. J. W. Jalowka, and T. E. Daubert, *Ind. Eng. Chem. Process Des. Dev.* **25**, 139 (1986).
 198. T. E. Daubert, and R. Bartakovits, *Ind. Eng. Chem. Res.* **28**, 5 (1989).
 199. S. Sugden, *The Parachor and Valency*, Routledge & Sons, London, 1930.
 200. H. P. Meissner, *Chem. Eng. Prog.* **45**, 2, 149 (1949).
 201. Y. Ogata, and M. Tsuchida, *Ind. Eng. Chem.* **49**, 415 (1957).
 202. W. Y. Lai, D. H. Chen, and R. N. Maddox, *Ind. Eng. Chem. Res.* **26**, 1072 (1987).
 203. H. S. Elbro, A. Fredenslund, and P. Rasmussen, *Ind. Eng. Chem. Res.* **30**, 2576 (1991).
 204. G. Le Bas, *The Molecular Volumes of Liquid Chemical Compounds*, Longmans, Green, New York, 1915.
 205. M. T. Tyn, and W. F. Calus, *J. Chem. Eng. Data* **20**, 106 (1975).
 206. S. Banarjee, P. H. Howard, and S. Lande, *Chemosphere* **21**, 11–12, 1173–1180, (1990).
 207. P. P. Radecki, T. N. Rogers, and M. E. Mullins, *Initial Development of a Method to Predict Antoine Constant Group Contributions by Simultaneous Optimization*, AIChE-DIPPR Project 912 Meeting, Michigan Technological University (June 8, 1992).
 208. C. Antoine, *Compt. Rend.* **107**, 681, 836 (1888).
 209. L. P. Burkhard, *Ind. Eng. Chem. Fundam.* **24**, 119 (1985).
 210. A. B. Macknick, and J. M. Prausnitz, *Ind. Eng. Chem. Fundam.* **18**, 348 (1979).
 211. A. Bondi, *Physical Properties of Molecular Crystals, Liquids and Gases*, John Wiley & Sons, Inc., New York, 1968, Chapt. 14.
 212. C. M. Kelly, P. M. Mathias, and F. K. Schweighardt, *Ind. Eng. Chem. Res.* **27**, 1732 (1988).
 213. D. D. Lawson, *Appl. Energy* **6**, 241 (1980).
 214. D. Hoshino, X. Zhu, K. Nagahama, and M. Hirata, *Ind. Eng. Chem. Fundam.* **24**, 112 (1985).
 215. D. Hoshino, K. Nagahama, and M. Hirata, *J. Chem. Eng. Japan* **11**, 403 (1978).
 216. S. W. Benson, *Thermochemical Kinetics*, John Wiley & Sons, Inc., New York, 1968, Chapt. 2.
 217. K. Ogiwara, Y. Arai, and S. Saito, *J. Chem. Eng. Japan* **14**, 156 (1981); Y. Yoneda, *Bull. Chem. Soc. Japan* **52**, 1297 (1979).
 218. T-P Thinh, J-L Duran, and R. S. Ramalho, *Ind. Eng. Chem. Process Des. Dev.* **10**, 576 (1971).
 219. K. Ogiwara, Y. Arai, and S. Saito, *J. Chem. Eng. (Japan)* **14**, 156 (1981).
 220. D. B. MacLeod, *Trans. Faraday Soc.* **19**, 38 (1923).
 221. D. Reichenberg, *Symp. Transp. Prop. Fluids and Fluid Mixtures*, Natl. Eng. Lab., East Kilbride, Glasgow, Scotland, 1979; also *AIChE. J.* **19**, 854 (1973); **21**, 181 (1975).
 222. D. van Velzen, R. L. Cardozo, and H. Langenkamp, *Ind. Eng. Chem. Fundam.* **11**, 20 (1972).
 223. T. E. Daubert, private communication, Department of Chemical Engineering, The Pennsylvania State University, University Park, Pa., 1992.
 224. D. Roy, and G. Thodos, *Ind. Eng. Chem. Fundam.* **7**, 529 (1968); **9**, 71 (1970).
 225. M. Nagvekar, and T. E. Daubert, *Ind. Eng. Chem. Res.* **26**, 1362 (1987).
 226. D. Hoshino, K. Nagahama, and M. Hirata, *J. Chem. Eng. Japan* **15**, 153 (1982).
 227. M. E. Mullins, T. N. Rogers, A. J. Pintar and J. R. Mihelcic, *AIChE Design Institute for Physical Property Data, Research Project 912, Environmental, Safety and Health Data Estimation Manual*, Year End Progress Report, Michigan Technological University, Houghton, Mich. March, 1992, pp. 91–114.
 228. T. N. Rogers, *Henry's Law Constants for Environmental Applications*, presented at the Second International Conference on Thermodynamics of Aqueous Systems with Industrial Applications, Warrenton, Virginia (May 10–14, 1987).
 229. W. B. Arbuckle, *Environ. Sci. Technol.* **17**, 537 (1983).
 230. C. Antunes and D. Tassios, *Ind. Eng. Chem. Process Des. Dev.* **22**, 457 (1983).
 231. J. Hine and P. K. Mookerjee, *J. Org. Chem.* **40**, 292 (1975).
 232. W. J. Lyman, Estimation of Physical Properties in W. B. Neely and G. E. Blau, *Environmental Exposure from Chemicals*, Vol 1, CRC Press, Boca Raton, Fla., 1987, pp. 13–47.

233. W. M. Meylan, and P. H. Howard, *Env. Toxicology and Chem.* **10**, 1283 (1991).

234. G. G. Nys, and R. F. Rekker, *Eur. J. Med. Chem. Chim. Ther.* **8**, 521 (1973); **9**, 361 (1975).

235. A. Leo, P. Y. C. Jow, C. Silipo, and C. Hansch, *J. Med. Chem.* **18**, 865 (1975).

236. J. T. Chou, and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **19**, 172 (1979).

237. E. N. Fuller, P. D. Schettler, and J. C. Giddings, *Ind. Eng. Chem.* **58**, 19 (1966).

238. D. P. Northup, A. J. Engel, R. P. Danner, and T. E. Daubert, *Design Institute for Physical Property Data Documentation Report No. 802-10-86*, American Institute of Chemical Engineers, New York, 1987.

239. M. T. Tyn and W. F. Calus, *J. Chem. Eng. Data* **20**, 106 (1975).

240. W. Hayduk and B. S. Minhas, *Can. J. Chem. Eng.* **60**, 295 (1982).

241. J. A. Abusleme and J. H. Vera, *AIChE. J.* **35**, 481 (1989).

242. S. W. Karickhoff, V. K. McDaniel, C. Melton, A. N. Vellino, D. E. Nute, and L. A. Carrieria, *Env. Toxicology and Chem.* **10**, 1405 (1991).

243. J. P. Hickey, A. Aldridge, D. R. May Passino, A. M. Frank, *Expert System Predicts Aquatic Toxicity from Contaminant Chemical Structure*, National Fisheries Research Center-Great Lakes, U.S. Fish and Wildlife Service, Ann Arbor, Mich., 1991; *Ibid.*, *Drug Information Journal* **26**, 487 (1992).

244. K. L. E. Kaiser, ed., *Proceedings of the Workshop on Quantitative Structure-Activity Relationships (QSAR) in Environmental Toxicology held at McMaster University, Hamilton, Ontario, Canada, August 16-18, 1983*, D. Reidel Publishing Co., Dordrecht, Germany, 1983.

245. J. D. McKinney, *Environmental Health Perspectives* **61**, 5 (1985).

246. C. M. Auer, J. V. Nabholz, and K. P. Baetke, *Environmental Health Perspectives* **87**, 183 (1990).

247. P. N. Craig and K. Enslein, *Structure-Activity in Hazard Assessment in Hazard Assessment of Chemicals Current Developments*, Vol. 1, Academic Press, New York, 1981, p. 381.

248. S. P. Gupte, *Chem. Rev.* **89**, 1765 (1989).

249. J. C. Dearden, *Environmental Health Perspectives* **61**, 203-228 (1985).

250. M. L. Tosato, L. Viganò, B. Skagerberg, and S. Clementi, *Environ. Sci. Technol.* **25**(4), 695 (1991).

251. J. M. McKim, S. P. Bradbury, and G. J. Niemi, *Environmental Health Perspectives* **71**, 171 (1987).

252. J. L. M. Hermans, *Environmental Health Perspectives* **87**, 219 (1990).

253. J. R. Coats, *Environmental Health Perspectives* **87**, 255 (1990).

254. L. P. Hammett, *Physical Organic Chemistry*, 1st Ed., McGraw-Hill, New York, 1940, p. 184.

255. M. Charton, *Chemtech* **4**, 502 (1974).

256. C. Hansch, J. E. Quinlan, and G. L. Lawrence, *J. Org. Chem.* **33**, 347 (1968).

257. R. W. Taft, J.-L. M. Abboud, M. L. Kamlet, and M. H. Abraham, *J. of Solution Chem.* **14**, 3, 153 (1985). An excellent review source including an extensive list of properties correlated with solvatochromic parameters.

258. J. P. Hickey, and D. R. Passino-Reader, *Environ. Sci. Technol.* **25**, 1754 (1991).

259. M. H. Abraham, R. M. Doherty, M. J. Kamlet, and R. W. Taft, *Chemistry in Britain*, 551 (June, 1986); M. J. Kamlet, R. M. Doherty, J.-L. M. Abboud, M. H. Abraham, and R. W. Taft, *Chemtech*, 566 (Sept 1986).

260. D. C. Leuhrs, R. E. Brown, and K. A. Godbole, *J. of Solution Chem.* **18**(5), 463 (1989).

261. D. C. Leuhrs, and K. A. Godbole, *J. of Solution Chem.* **19**2, 187 (1990).

262. A. Leo, C. Hansch, and D. Elkins, *Chemical Reviews*, 71, 6, 525-616 (1971).

263. M. J. Hait, C. L. Liotta, C. A. Eckert, D. L. Bergmann, A. M. Karachewski, A. J. Dallas, D. I. Eikens, J. J. Li, P. W. Carr, R. B. Poe, and S. C. Rutan, *The SPACE Predictor for Infinite Dilution Activity Coefficients*, unpublished, presented at AIChE 1992 Annual Meeting, Nov. 3, Miami Beach, Fla. For information, write to Charles. A. Eckert, School of Chemical Engineering, Georgia Institute of Technology, Atlanta, Ga. 30332-0100.

264. D. Hawker, *Chemosphere* **19**, 1585-1593 (1989).

265. D. Hawker, *Chemosphere* **20**, 467-477 (1990).

266. Y. Marcus, *J. Phys. Chem.* **95**, 8886-8891 (1991).

267. M. J. Kamlet, R. M. Doherty, M. H. Abraham, and R. W. Taft, *Carbon* **23**(5), 549 (1985).

268. M. J. Kamlet, R. M. Doherty, M. H. Abraham, Y. Marcus, and R. W. Taft, *J. Phys. Chem.* **92**, 5244-5255 (1988).

269. H. Wiener, *J. Am. Chem. Soc.* **69**, 17 (1947).

270. H. Weiner, *J. Chem. Phys.* **15**, 766 (1947).

271. P. G. Seybold, M. May, and U. A. Bagal, *J. of Chemical Education* **64**, 575 (1987).

272. M. Randic, *J. Am. Chem. Soc.* **97**, 6609 (1975).

273. L. B. Kier, and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.

274. L. B. Kier, W. J. Murray, M. Randic, and L. H. Hall, *J. Pharm. Sci.* **65**, 1226 (1976).

275. L. B. Kier and L. H. Hall, *J. Pharm. Sci.* **65**, 1806 (1975).

276. L. H. Hall, L. B. Kier, and W. J. Murray, *J. Pharm. Sci.* **64**, 1974 (1975).

277. W. J. Murray, L. H. Hall, and L. B. Kier, *J. Pharm. Sci.* **64**, 1978 (1975).

278. L. B. Kier, *J. Pharm. Sci.* **70**, 930 (1981).

279. D. T. Stanton and P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **32**, 109 (1992).

280. M. B. Perry and C. M. White, *AIChE. J.* **33**, 146 (1987).

281. D. Weininger and J. L. Weininger, in C. Hansch, P. G. Sammes, and J. B. Taylor, eds., *Comprehensive Medicinal Chemistry, Vol. 4*, Pergamon Press, New York, 1990.

282. S. Borman, *C & E. News*, 28 (Sept. 18, 1989).

283. H. Harary, *Graph Theory*, Addison-Wesley, Reading, Mass., 1969.

284. R. J. Wilson, *Introduction to Graph Theory*, Academic Press, New York, 1972.

285. Z. Mihalic and N. Trinajstic, *J. of Chemical Education* **69**, 701 (1992).

286. H. P. Schultz, E. B. Schultz, and T. P. Schultz, *J. Chem. Inf. Comput. Sci.* **32**, 69 (1992).

287. K. Yuta, and P. C. Jurs, *J. Med. Chem.* **24**, 241 (1981).

288. M. N. Hasan, and P. C. Jurs, *Anal. Chem.* **55**, 263 (1983).

289. J. S. Wishnok In R. A. Scanlan and S. R. Tannenbaum, Eds., *N-Nitroso Compounds*, ACS Symposium Series, No. 174, Am. Chem. Soc., Washington, D.C., 1981, pp. 77-88.

290. R. Dagani, *Chem. Eng. News* **26**, (Mar 9, 1981).

291. K. Enslein, and P. N. Craig, *J. Toxicol. Environ. Health* **10**, 521 (1982).

292. D. H. Rouvray, *Am. Sci.* **61**, 729 (1973).

Michael E. Mullins
Tony N. Rogers
Michigan Technological University

Peter P. Radecki

Center for Clean Industrial and Treatment Technologies (CenCITT)

ENGINEERING PLASTICS

Engineering plastics account for 7^o % of all plastics production and about 9^o % of the thermoplastics market (1). Engineering plastics are of several generic types and, considering the rapid advances in polymerization, blending, alloying, filling, reinforcing, and reactive compounding technologies, the boundaries of their definitions have become increasingly blurred. For the purposes of this article, the following descriptive definitions are combined. Engineering plastics are those "plastics which lend themselves to use for engineering design, such as gears and structural members" (2). This implies that engineering plastics can substitute for traditional materials of construction, particularly metals. The third edition of the *Encyclopedia* defines engineering plastics as thermoplastic resins, neat or unreinforced or filled, which maintain dimensional stability and most mechanical properties above 100°C and below 0°C. This definition encompasses plastics that can be formed into functional parts that can bear loads and withstand abuse in temperature environments commonly tolerated by the traditional engineering materials: wood, metals, glass, and ceramics. The terms engineering plastics and engineering thermoplastics can be used interchangeably (3).

Commonly accepted practice restricts the term to plastics that serve engineering purposes and can be processed and reprocessed by injection and extrusion methods. This excludes the so-called specialty plastics, eg, fluorocarbon polymers and infusible film products such as Kapton and Upilex polyimide film, and thermosets including phenolics, epoxies, urea–formaldehydes, and silicones, some of which have been termed engineering plastics by other authors (4) (see ELASTOMERS, SYNTHETIC FLUOROCARBON ELASTOMERS; FLUORINE COMPOUNDS, ORGANIC TETRAFLUOROETHYLENE COPOLYMERS WITH ETHYLENE; PHENOLIC RESINS; EPOXY RESINS; AMINO RESINS AND PLASTICS).

Many other polymers can be processed by injection molding and extrusion, but do not exhibit enough toughness, temperature resistance, or dimensional stability for engineering use. Generic resins falling within the scope of this definition include acetal resins (qv), polyamides (qv) (nylons), polyimides (qv), polyetherimides, polyesters (qv), liquid crystal polymers, and selected polyolefins, blends, or alloys of the foregoing resins, and some examples from other resin types (see also POLYETHERS). Thermoplastic elastomers and commodity resins, such as most styrenic resins, acrylics, polyolefins, polyurethanes, poly(vinyl chloride)s, and related chlorinated polyolefins, lose mechanical properties below 100°C (see also ELASTOMERS, SYNTHETIC THERMOPLASTIC ELASTOMERS). Acrylonitrile–butadiene–styrene polymers (ABS) maintain engineering properties below 100°C and are excluded from engineering resins by the authors. Unreinforced nylons exhibit some creep under load but are included in the engineering plastics family, as opposed to reinforced polypropylene, whose properties are mainly a function of high reinforcing fiber bonding (ie, Azdel) rather than of the resin itself. As noted previously, a precise definition of engineering plastics is elusive and inclusion or exclusion of a resin from the family is often arbitrary.

The development of engineering thermoplastic resins (4,5) began with the pioneering work of Carothers on nylons at E.I. du Pont de Nemours & Co., Inc. in the 1930s (6) continued with General Electric Co.'s introduction of such resins as polyetherimide Ultem resin (7), and more recently high temperature polyolefins such as polycyclopentanes, its copolymers, and polymethylpentane have appeared. Certain trends are discernible. First, the introduction of new polymers has slowed. Several new varieties of established polymer resins were introduced in the 1960s and 1970s, but few new generic polymers have been commercialized since the 1980s. This is not surprising in light of the high risk associated with the introduction of new materials which cost millions of dollars and years of development. On the other hand, many new blends and alloys of engineering materials, often tailored to specific applications, have been introduced in the 1990s. Although their development often was the result of highly sophisticated and creative science, their commercialization typically required only 2–4 years. Plant and equipment investment costs were minimal since compounding and blending equipment is used for both development and manufacture. This success is evidenced by 6–8^o % growth rates for special blends of commodity polyolefins.

Plastics are either crystalline or amorphous (8). Acetals, most nylons, some polyesters, poly(aromatic sulfides), and polyetherketones are crystalline resins. Nonoriented crystalline polymers do not transmit light, whereas amorphous polymers have some degree of transparency. The mode of processing or treatment, eg, biaxial orientation, can alter the morphological form. However, a polymer is considered crystalline if that is the morphology which develops upon nonaccelerating cooling of the melt to ambient temperature. Amorphous polymers are defined similarly. When treatment of polymers results in morphological change, properties can change as well. For example, poly(ethylene terephthalate) (PET) when allowed to slowly solidify from a melt, becomes a partially crystalline, opaque solid. If it is cooled rapidly from the melt (quenched), however, the solidified material is amorphous and transparent.

Besides light transmission, there are other properties distinguishing amorphous from crystalline polymers (Table 1). Organic solvent resistance tends to be higher for crystalline systems. In applications where solvent resistance is important, such as in a gear box exposed to oils and greases, a crystalline polymer is a likely choice. On the other hand, for filter bowls, glazing, etc, where clarity is required, an amorphous polymer might be preferred.

Table 1. Morphology Effects and Polymer Properties

Property	Crystalline	Amorphous
organic solvent resistance	high	low
light transmission	none to low (random)	high
lubricity	high	low
dimensional stability	high ^a	moderate
mold shrinkage	high	low

^a With a high degree of crystallinity and no water absorption.

Lubricity of crystalline polymers is usually higher than that of amorphous polymers. Excellent machinery parts are made from crystalline nylon-6,6 resins, eg, gears, cams, wedges, and other components not requiring lubrication. Gears made of amorphous polyimide resin, on the other hand, do not exhibit this feature.

Crystalline polymers tend to be more stable dimensionally above their glass-transition temperature than amorphous materials. The closer packing and regular physical interactions of the crystalline species would be expected to yield less creep and susceptibility to deformation. Crystalline nylons expand and contract with changes in ambient moisture (9). Thus changes in humidity can result in warpage. This is important for large parts where dimensional change of a few micrometers per centimeter can result in serious deformation. Close packing of crystalline polymer molecules further causes them to exhibit high mold shrinkage compared with amorphous materials. Mold shrinkages of most amorphous resins are ca 0.005 mm mm, whereas crystalline resins shrink ca 0.015 mm mm (10).

Polymers with differing morphologies respond differently to fillers (qv) and reinforcements. In crystalline resins, heat distortion temperature (HDT) increases as the aspect ratio and amount of filler and reinforcement are increased. In fact, glass reinforcement can result in the HDT approaching the melting point. Amorphous polymers are much less affected. Addition of fillers, however, interrupts amorphous polymer molecules' physical interactions, and certain properties, such as impact strength, are reduced.

Engineering thermoplastics are priced between the very expensive resins and the high volume, low priced commodities, eg, poly(phenylene oxide)–polystyrene alloys at \$2 /kg to polyetheretherketones at \$60 /kg (Fig. 1). Many specialty and commodity resins are addition polymers. Their prices are primarily governed by raw material costs and the complexity of the manufacturing process. Thus the raw material for fluoropolymers is much more expensive than the olefins used for polyethylenes (PE), polystyrenes (PS), acrylics, vinyl chloride polymers, etc. Engineering polymers tend to be composed of aromatic monomers, except for acetals and nylons, and their monomers are usually linked by condensation (ester, carbonate, amide, imide), substitution (sulfide) or oxidative coupling (ether), and in a few cases, addition (polyolefins). High monomer costs and complex polymerization processes force higher prices (12). Larger volumes of production take advantage of economy of scale. Representative producers and trademarks of engineering plastics are given in Table 2. Key producers include multinational companies, with plants in many countries. Recent political events such as the opening of Eastern Europe, the North American Free Trade Zone, the advancing European Economic Union, and sharpening competition are increasing the multinational efforts to seek local pricing advantages by use of cheaper local labor, elimination of transport charges, and local tax incentives.

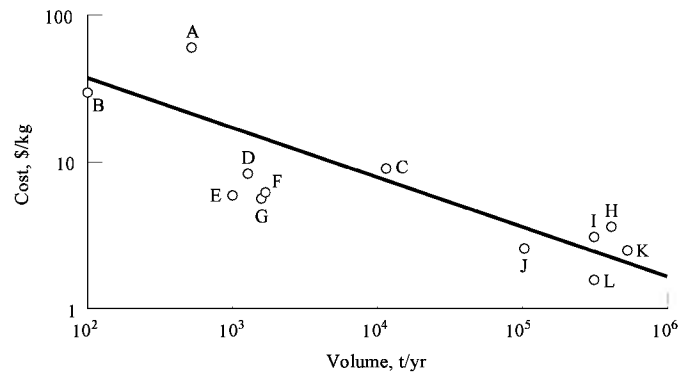


Fig. 1. Engineering resins cost vs annual volume (11) (HDT, °C): A, polyetheretherketone (288); B, polyamideimide (>270); C, polyarylether sulfone (170– >200); D, polyimide (190); E, amorphous nylons (124); F, poly(phenylene sulfide) (>260); G, polyarylates (170); H, crystalline nylons (90–220); I, polycarbonate (130); J, midrange poly(phenylene oxide) alloy (107–150); K, polyphthalate esters (180–260); and L, acetal resins (110–140).

Table 2. U.S. Engineering Resins Suppliers

Company	Trade name	Resin
Allied-Signal	Capron	polyamide
	Nypel	polyamide
	Petra	poly(ethylene terephthalate)
	Dimension	blend
Amoco Performance Polymers	Xydar	liquid crystal polymer
	Torlon	polyamideimide
	Amodel	polyphthalamide
	Radel	polyarylsulfone
	Radel	polyphenylenesulfone
	Ardel	polyarylate
	Mindel	polysulfone
	Udel	polysulfone
	Kadel	polyketone
	Ultramid	polyamide
BASF	Ultradur	poly(butylene terephthalate)
	Ultrason	polyethersulfone
	Ultraform	acetal
	Ultrason	polysulfone
	Ultrapek	polyaryletherketone
	Ultrablend	blend
	Calibre	polycarbonate
Dow	Pulse	blend
	Stanyl	polyamide
DSM	Delrin	acetal
Du Pont	Zytel	polyamide
	Minlon	polyamide
	Rynite	poly(butylene terephthalate)
	Rynite	poly(ethylene terephthalate)
	Kodapak	poly(ethylene terephthalate)
Eastman	Ektar	poly(ethylene terephthalate)
	Kodak	PETG
	Grilamid	polyamide
EMS–American Grilon	Grilon	polyamide
	Valox	poly(butylene terephthalate)
GE Plastics	Valox	poly(ethylene terephthalate)
	Valox	poly(cyclohexylene terephthalate)
	Lexan	polycarbonate
	Xenoy	blend
	Ultem	polyetherimide
	Noryl	poly(phenylene oxide)–polystyrene
	Supec	poly(phenylene sulfide)
	Celcon	acetal
Hoechst-Celanese	Celstran	long-strand fibers
	Celanex	poly(butylene terephthalate)
	Impet	poly(ethylene terephthalate)
	Vandar	blend
	Celanese Nylon	polyamide

Hbbs	Fortron	poly(phenylene sulfide)
	Vectra	liquid crystal polymer
	Celazole	polybenzimidazole
	Vestamid	polyamide
	Vestodur	poly(butylene terephthalate)
ICI Advanced Materials	Vestoran	poly(phenylene ether)
	Victrex	polyetheretherketone
	Maranyl	polyamide
Miles (formerly Mobay)	Nydur	polyamide
	Makrolon	polycarbonate
	Apec	polycarbonate
	Makroblend	blend
	Bayblend	blend
Monsanto	Vydyne	polyamide
	Triax	blend
Phillips	Ryton	poly(phenylene sulfide)
Rhone Poulenc	Technyl	polyamide

Processing

Injection Molding. This is the process most commonly used for engineering thermoplastics (13) (see POLYMER PROCESSING). Typically, the resin is processed as a pellet, ca 2 mm in diameter and 3 mm long, made by chopping an extruded strand, and drying in a forced-draft oven. The pellets are introduced into a hopper which feeds into a screw- or plunger-actuated melt chamber. The melted plastic is injected into a mold in which it solidifies to the finished article. The size and complexity of the part capable of being produced are determined by the flow characteristics of the molten resin at the processing temperature. Typically, the viscosity of a resin should be ca 10,000 Pa·s (1 × 10⁵ P) at a shear rate of 100 s⁻¹ for convenient injection molding or extrusion.

Foam-injection molding is a variation of conventional injection molding (15,16). The molten resin, under pressure in the extruder melt chamber, is injected with an inert gas or an *in situ* heat-activated chemical blowing agent which decomposes to produce gas. When the resin is injected into the mold, it leaves the high pressure melt chamber and enters the lower pressure mold cavity. The hot gases expand to form a honeycomb structure in the solidified foam plastic part. Larger parts can be produced by foam molding than by conventional injection molding. Parts are internally reinforced by the honeycomb structure which also results in thick part walls.

Blow Molding. In blow molding, the molten resin is extruded as a hollow, cylindrical parison (17–19). The mold shell is closed over the parison which is blown against the mold, resulting in a hollow part. The resin should exhibit melt strength under the low shear conditions on the parison but good flow during the high shear blowing process. Some engineering resins are specifically supplied for blow molding, eg, slightly branched polycarbonate (PC) resin by Bayer AG, The Dow Chemical Company, and the General Electric Company.

Blown film is made similarly to blow-molded articles; the molten resin is extruded continuously as a parison, which is blown to form a continuous film cylinder. This cylinder can be rolled up as a two-layer structure or slit to form one large or two smaller single layers.

Extrusion. Sheet, film, and profiled articles are made by extrusion (20). The resin is melted and forced through a die plate or head. Variations include multilayer and blown film applications. In multilayer coextrusions, different combinations of plastics are separately but concurrently extruded to form layered sheet or film. In the packaging industry, specialty resins such as high barrier ethylene–vinyl alcohol copolymers are combined with heat- and impact-resistant thermoplastics for food packages. The properties of each resin layer are additive, as opposed to the "averaging of property" in blends. Multilayers are also used for blow-molded containers, films, and sheet products (see also FILM AND SHEETING MATERIALS).

Thermoforming of extruded sheet is achieved by warming the sheet to its glass-transition point and collapsing it over a mold shape (21). The collapse can be effected by vacuum suction or pressures applied on the sheet above the mold. This technique is used to make trays and profiled advertising signs. Both single- and multilayer sheets can be used.

Extruded engineering thermoplastic stock can be treated like other building material in that it can be machined, cut, and fastened. However, none of the engineering plastics can be considered a one-for-one substitute for metals or wood. For example, impact resistance must be considered, and glues, paints, etc, must be screened for chemical aggressiveness and adhesion capability.

Reaction-Injection Molding and Reactive Casting. Reaction-injection molding (RIM) (22) and reactive casting (23) have been demonstrated on nylon-6, which is polymerized by catalytic ring opening and linear recondensation of ε-caprolactam (qv) (24).

Another example is provided by cyclopolyolefins. Dicyclopentadiene (DCPD) is polymerized via a RIM process using tungsten-based catalysts into a three-dimensional network containing bicyclic rings connected by double bonds. The material was introduced by Hercules in 1985 and called Metton (25). In the process, two separate liquid streams are mixed and allowed to react in the mold. Both streams are DCPD, but one contains one portion of the catalyst (a Lewis base) the other contains the remaining portion (tungsten-based). The resultant resin has properties that put it in the class of engineering resins and causes it to have a much higher price than typical commodity polyolefins (see CYCLOPENTADIENE AND DICYCLOPENTADIENE).

Engineering thermoplastics have also been used in preimpregnated constructions. The thermoplastic is thoroughly dispersed as a continuous phase in glass, other resins, carbon fibers (qv), or other reinforcement. Articles can be produced from these constructions using thermoforming techniques. For example, the aerospace industry uses polyetheretherketone (PEEK) in woven carbon-fiber tapes (26). Experimental uses of other composite constructions have been reported (27) (see also COMPOSITE MATERIALS, POLYMER MATRIX).

Properties

For physical, thermal, electrical, and mechanical properties, ASTM test methods are employed (28). Flammability ratings are often based on Underwriters Laboratories' (UL) standards (29). UL flammability ratings given in this article are not intended to reflect the hazards presented by the resins under use conditions. Typical properties are given in Table 3. More details and additional properties are given in References 5 and 31–33.

Table 3. Units of Resin Properties^a

Property	ASTM method	Unit	
		English	SI
Physical			
density (specific gravity)	D792		g cm ³
water absorption, 24 h at 23°C	D570	wt %	wt %
mold shrinkage	D955	mil in.	mm mm
transmittance	D1003	%	%
haze	D1003	%	%
refractive index		dimensionless	
yellowness index (YI)	D1925	dimensionless	
Thermal			

deflection temperature	D648		
at 0.46 N mm ^{2b}		°F	°C
at 1.82 N mm ^{2b}		°F	°C
specific heat		Btu (lb·°F)	J (kg·K)
thermal conductivity		Btu·in. (h·ft ² ·°F)	W (m·K)
coefficient of thermal expansion	D696	per °F	per °C
flammability ratings ^c UL Standard 94			
1.6-mm thickness			various codes
3.2-mm thickness			various codes
	<i>Electrical</i>		
dielectric strength, short time, 3.2-mm thickness	D149	V mil	kV mm
dielectric constant	D150		
60 Hz			dimensionless
10 Hz			dimensionless
dissipation factor	D150		
60 Hz			dimensionless
10 Hz			dimensionless
volume resistivity at 23°C, dry	D150	Ω·cm	Ω·cm
	<i>Mechanical</i>		
tensile strength	D638		
yield		psi	N mm ^{2b}
ultimate		psi	N mm ^{2b}
elongation	D638	°	°
flexural strength	D790	psi	N mm ^{2b}
flexural modulus	D790	psi	N mm ^{2b}
shear strength	D732		
yield		psi	N mm ^{2b}
ultimate		psi	N mm ^{2b}
Izod impact strength, 3.2-mm thickness	D256		
notched		ftlbf in.	J m
unnotched		ftlbf in.	J m
tensile impact strength, S	D1822	ftlbf in. ²	kJ m ²
falling-dart impact strength, 3.2-mm thickness		ftlbf	J
Rockwell hardness	D785		
M			dimensionless
R			dimensionless
	<i>Rheological</i>		
viscosity			
dynamic		poise	Pa·s
intrinsic, [η]			dL g
kinematic			mm ² s
spiral flow, 3.2-mm thickness		in.	cm
melt index	D1238		g

^a Ref. 30.

^b N mm² MPa.

^c These ratings are not intended to reflect the hazards presented by any material under actual use conditions.

Physical Properties. Physical properties include specific gravity, water absorption, mold shrinkage, transmittance, haze, and refractive index. Specific gravity affects performance and has commercial implications. The price of the material divided by the specific gravity gives the yield in cost per unit volume. Comparison of yields gives an evaluation of raw material costs.

Mold Shrinkage. Mold shrinkage of unreinforced crystalline plastics is high. This is critical in part and mold design, especially for large parts.

Water Absorption. If water is absorbed, predrying is required before injection molding or extrusion. At high processing temperatures, most absorbed water is converted to superheated steam. Failure to remove absorbed water may result in degradation of the material and poor performance. For example, molded poly(ethylene terephthalate) (PET) containing 0.1° water exhibits splay marks on the surface. In polycarbonate (PC), 0.1° water causes backbone degradation. Extrusion at 290°C of undried linear PC with an intrinsic viscosity [η] of 0.69 dL g reduces [η] to 0.50 dL g. Absorbed water can thus greatly affect final properties.

In nylons, the amide linkages accept water molecules through hydrogen-bonding interactions. The absorbed water content varies with humidity, and marked changes in properties are observed. For example, the volume electrical resistivity of nylon-6 is reduced by a factor of 10⁷ between anhydrous and 100° rh conditions (34); mechanical property changes occur also. When dried at 23°C, the notched Izod impact strength is 40 J m (0.75 ftlbf in.). With 3° water absorption, the impact increases to 150 J m (2.8 ftlbf in.) (35). Water absorption after molding can result in warpage. Thus plastics that tend to absorb water are not suitable for making large parts which must maintain dimensional stability. Dimensional changes can be minimized by compounding or blending with mineral fillers or reinforcing fibers. Some producers offer amorphous polyamide resins which exhibit reduced moisture absorption.

Color/Transparency. Almost all amorphous engineering thermoplastics, except PC and some polyester carbonates, are inherently colored. Even polycarbonates have yellowness indexes (YI) (36) of 0.1 to 5.0. Colorless material is produced from these resins by compounding with complementary blue dyes which reduce transmission. Haze in amorphous resins is an indication of particulates. Haze reduces optical clarity and transmission.

Thermal Properties. Thermal properties include heat-deflection temperature (HDT), specific heat, continuous use temperature, thermal conductivity, coefficient of thermal expansion, and flammability ratings. Heat-deflection temperature is a measure of the minimum temperature that results in a specified deformation of a plastic beam under loads of 1.82 or 0.46 N mm² (264 or 67 psi, respectively). For an unreinforced plastic, this is typically ca 20°C below the glass-transition temperature, *T_g*, at which the molecular mobility is altered. Sometimes confused with HDT is the UL Thermal Index, which Underwriters Laboratories established as a safe continuous operation temperature for apparatus made of plastics (37). Typically, UL temperature indexes are significantly lower than HDTs. Specific heat and thermal conductivity relate to insulating properties. The coefficient of thermal expansion is an important component of mold shrinkage and must be considered when designing composite structures.

The UL flammability ratings describe the relative ease of ignition and combustibility of plastics. Tests include the measurement of flame propagation, time to self-extinguish, melt and drip with and without flame, and oxygen indexes. Some engineering plastics, eg, polyetherimides, are, as ranked by this test, inherently nonflammable. Others can be made nonflammable by compounding with flame retardants (FRs) such as bromine

compounds, antimony oxide, or phosphates. No valid statements about flammability can be made without knowledge of FR ingredients relative to the polymer being flame retarded. The UL ratings of each product must be considered.

Electrical and Mechanical Properties. Electrical properties include dielectric strength, dielectric constant, dissipation factor, and volume resistivity; these properties can change with temperature and absorbed water.

For a part to exhibit structural stiffness, flexural moduli should be above 2000 N/mm² (290,000 psi). Notched Izod impact values should be determined at different thicknesses. Some plastics exhibit different notch sensitivities. For example, PC, 3.2 mm thick, has a notched Izod impact of 800 J/m (15 ft-lbf/in.) which drops to 100 J/m (1.9 ft-lbf/in.) at 6.4-mm thickness. On the other hand, one bisphenol A phthalate-based polyarylate resin maintains a 250-J/m (4.7-ft-lbf/in.) notched Izod impact at both thicknesses. Toughness depends on the structure of the part under consideration as well as the plastic employed to make the part. Mechanical properties, like electrical properties, are also subject to thermal and water-content changes.

Chemical Resistance. Chemical resistance is less rigidly defined than the previously discussed properties. Measurement methods include immersion in selected vapors or liquid and response to various surface treatments, including painting. Elevated temperature, temperature cycling, wet-dry cycling, application of stress and post-treatments are often employed. Parts tested are typically ASTM flexural, tensile, or impact specimens, or a commercial part. Resistance to chemical treatment is rated by comparing mechanical and optical properties before and after stress by organic chemicals; aqueous acids, bases, salts, and buffers; light of various wavelengths; or combinations of the foregoing agents.

In many commercial brochures, chemical resistance is indicated as excellent, good, fair, or poor. Although the test method is usually outlined, wide interpretation is possible. Immersion tests are usually described in this manner. Hydrolytic stability is tested by salt-spray cycling or autoclave cycling.

A common semiquantitative chemical immersion test incorporates stress and uses Bergen strain jigs (38). An ASTM flex bar is bent over a shallow arc of metal during the immersion. The arc varies from almost no bend on one end to pronounced curvature at the other. The stress on the part therefore increases along the arc. Crazing or cracking indicates failure. Results are reported in terms of time to failure and the distance along the arc where failure initiates. This method is used when testing resistance to gasoline, oils, brake fluids, and cleaning solvents.

In a clear amorphous plastic, effects on light transmission are easily measured. Retention of surface gloss or mechanical properties such as impact can also be used. Susceptibility to photochemical factors is important for glazing and outdoor applications (39). Many engineering plastics contain aromatic molecules and absorb uv and visible light, particularly light with wavelengths between 2500 and 3500 nm. Laboratory tests use artificial light sources; outdoor testing exposes the part to solar radiation. Optical, gloss, and mechanical property changes are used to measure effects.

Rheological Properties. Rheological properties of engineering thermoplastics are determined in both solid and molten states. Melt-flow behavior is important for the design of plastic parts and molds and affects formation of knit lines, locations of internal stresses, mold fill, melt strength, and the conditions necessary for extrusion and molding. Useful rheological information is gained from viscosity as a function of shear rate. For example, Figure 2 shows plots for two resins, A and B. Resin B exhibits less tendency to lose viscosity at shear rates above 100 s⁻¹ than does A. This suggests that A is less Newtonian than B in its flow properties and exhibits more melt strength at low shear and greater fluidity at high shear. Polymer A would therefore make a better candidate for blow molding than B because the molten parison would sag less after extrusion. On the other hand, B should be extruded and injection molded at lower shear with greater facility.

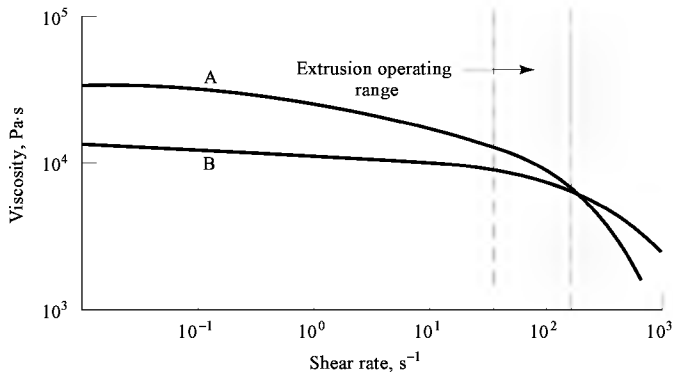


Fig. 2. Shear sensitivity of polycarbonate resins at 320°C (1 Pa·s = 10 P) (14). Polymers A and B are discussed in the text.

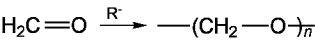
When a complete rheometric profile is not available, resin processibility can be estimated in a number of ways. Knowledge of the kind of plastic, its filler level, and intrinsic viscosity [η] is often sufficient for an experienced plastics technologist. Comparative melt-flow data are obtained with one-point extrusion rheometer measurements taken at a reference temperature at fixed process conditions. Data are usually based on the extrusion time or on the amount of plastic that is extruded in a given time. For example, using an extrusion plastometer (Tinius Olson Testing Machine Co.), the relative time to extrude a polycarbonate resin at 300°C vs a polyarylate resin is 6,000 vs 66,000 units. This suggests that the polyarylate is more viscous than PC at this temperature. Another test measures the spiral flow of resins by injection molding. The longer the resin travels within the spiral, the higher the melt flow. In the case of the PC vs polyarylate, the PC would be expected to have a much greater flow length than the polyarylate.

Complete property profiles including rheological data are not available for all grades of engineering thermoplastics. However, the situation is changing as a result of analytical efforts by the resin suppliers. These data are essential for CAD and CAE (computer-aided design and computer-aided engineering), which are being applied to engineering plastic part and mold designs (40,41) (see also RHEOLOGICAL MEASUREMENTS).

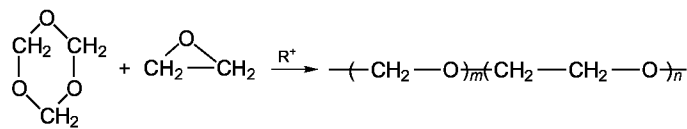
Mechanical Properties. Along with thermal properties, mechanical properties are key to the utility of engineering plastics. In the majority of applications the materials are subject to mechanical stress, often in conjunction with thermal stress. Stiffness and impact strength are among the most desirable properties of a plastic. The first allows for easy and economical design, coming closest to the classical design tradition based on metal properties; the second guarantees that the inevitable sudden stresses which everyday articles experience in everyday use will not destroy a given article or prematurely shorten its useful lifespan. Frequently, fillers such as glass and other fibers are compounded into plastics to increase their stiffness. Especially in the case of brittle resins this also results in an increase in the tensile and impact strength. To improve impact, special rubber additives of small particle size are employed. Typically these impact modifiers are produced by emulsion polymerization and they are almost always copolymers whose outermost layer is composed of a polymer designed to be compatible with the future host resin. Another approach frequently used in impact modification is the chemical reaction of an elastomer, eg, ethylene-propylene-diene, suitably modified to react with the amine functional groups of the polyamide molecule. Impact modifiers are believed to act by reducing formation of large cracks through craze formation, thereby absorbing energy.

Mainstream Engineering Resins

Acetal Resins. Acetal resins (qv) are poly(methylene oxide) or polyformaldehyde homopolymers and formaldehyde [50-00-0] copolymerized with aliphatic oxides such as ethylene oxide (42). The homopolymer resin polyoxymethylene [9002-81-7] (POM) is produced by the anionic catalytic polymerization of formaldehyde. For thermal stability, the resin is endcapped with an acyl or alkyl function.



The copolymers, eg, Celcon M 90 [95327-43-8], are produced by the cationic catalyzed reaction of 1,3,5-trioxane [110-88-3] and an ethylene oxide.



About six firms, not including their subsidiaries, make acetal resins; Hoechst-Celanese, Du Pont, and their subsidiaries are the principal producers. Polyacetals are translucent white and crystalline. They are solvent-resistant and self-lubricating. Unlike nylons, they have low moisture absorption and maintain dimensional stability under varying humidity. Melt-flow characteristics are excellent, allowing for short mold cycles. Good fatigue allows applications such as ski-binding parts. On the other hand, impact resistance is low. Mold design must allow for shrinkage above 25° 0 for unfilled resins. Specific gravities are high. Effective flame retardants have not been developed. For these reasons, applications are limited to small parts or those for which mold warpage, flammability, and toughness are not important. The choice of homopolymer vs copolymer resin involves some tradeoffs. For example, homopolymers have higher HDTs but lower unloaded maximum use temperatures and dielectric strengths. These could be important considerations for electrical and electronic applications. Processibility and thermal stability of the copolymer are the key advantages of the copolymer; there is less formaldehyde generation at higher processing temperatures. Properties are given in Table 4.

Table 4. Properties of Acetal Resins^a

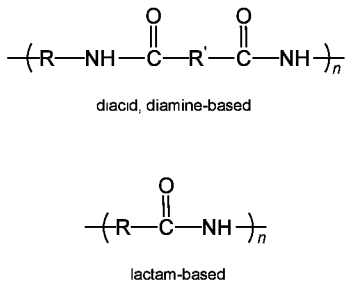
Property	ASTM method	Delrin 500 homopolymer resin ^b	Celcon M90 copolymer resin ^{c,d}
HDT at 1.82 N mm ^{2e} , °C	D698	136	110
notched Izod at 20°C, 3.2-mm thickness J m ^f	D256	75	70
flex modulus at 20°C, N mm ^{2e}	D790	2830	2625
specific gravity	D792	1.42	1.41
flammability rating, UL 94		HB	HB

^a Ref. 43.
^b Refs. 44, 45.
^c Oxymethylene and oxyethylene groups.
^d Ref. 46.
^e To convert N mm² to psi, multiply by 145.
^f To convert J m to ftlbf in., divide by 53.38.

Acetal resins are processed mainly by injection molding. Most resin products are virgin or pigmented neat resins; uv-light stabilized, glass-reinforced, impact modified, metal-plateable, and conductive grades are available. Applications are in consumer, appliance, transportation, and industrial and construction industries. Consumer articles include cassette cartridges, zippers, and container valves. Automotive parts include mechanical components such as cranks, gears, and bearings. Although the parts are small, the number of applications is large and a typical American automobile contains ca 1 kg of polyacetal. In the industrial and construction areas, valves and connectors are made of acetal resins. Because they are easily chrome-plated, polyacetals are often used in plumbing fixtures.

Worldwide acetal resin capacity is ca 228,000 t yr (47). Applications are tied to the consumer, appliance, construction, and automotive markets and will probably grow only as fast as the world economy. Growth rates of 2–10° 0 are projected for various areas (42).

Nylon Resins. Nylon engineering thermoplastic resins have the following polyamide structures:



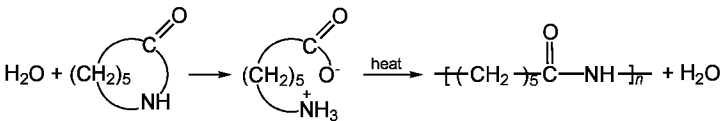
The straight-chain aliphatic nylons are commonly designated by the chain lengths of the monomer constituents. For example, the largest volume nylon, poly(hexamethylene adipamide) [32131-17-2], has hexamethylenediamine and adipic acid as monomer constituents and is called nylon-6,6; the diamine chain length is noted first. When a lactam is the precursor monomer as in polydodecanoamide [24937-16-4], its chain-length number is used, and the polymer is called nylon-12. When aromatic monomers are used, a letter is used as designation. Poly(hexamethylene isophthalamide) [25722-07-0] can be designated nylon-6,I. This system of nomenclature breaks down when a nylon is composed of three or more monomers. However, since most commercial nylons are made with one or two aliphatic monomers, the system has worked well in practice (see also POLYAMIDES).

Commercial engineering thermoplastic nylons are mainly crystalline resins. Nylon-6,6 [32131-17-2] is the largest volume resin, followed by nylon-6 (48). Other commercially available but much lower volume crystalline nylons are -6,9, -6,10, -6,12, -11, and -12. The crystallinity of the molded part decreases with chain size (49). A few truly amorphous commercial nylon resins contain both aromatic and aliphatic monomer constituents (50). For example, Trogamid T resin is made from a mixture of 2,2,4- and 2,4,4-trimethylhexamethylenediamines and terephthalic acid (51).

In the United States, 11 companies manufacture nylon resin; Du Pont, Monsanto, Hoechst-Celanese, and Allied-Signal are the leaders. In Europe, there are 22 manufacturers, and in Japan, six (52).

Nylon resins are made by numerous methods (53) ranging from ester amidation (54) to the Schotten-Baumann synthesis (55). The most commonly used method for making nylon-6,6 and related resins is the heat-induced condensation of monomeric salt complexes (56). In this process, stoichiometric amounts of diacid and diamine react in water to form salts. Water is removed and further heating converts the carboxylate functions to amide linkages. Chain lengths are controlled by small amounts of monofunctional reagents. The molten finished nylon resin can be directly extruded to pellets.

The preparation of nylon resins from lactam precursors involves ring opening, which is facilitated by a controlled amount of water in the reaction mixture. The salt complex condenses internally to produce the polyamide (57). The synthesis of nylon-6 [25038-54-4] from ε-caprolactam is as follows:



Physical properties of commercial nylons are shown in Table 5. Crystalline nylons are white and chemically and hydrolytically stable. Flammability, as measured by laboratory tests, varies with composition; the higher the aliphatic carbon content, the more flammable the resin. Both filled and unfilled nylon resins exhibit lubricity. Mineral- and glass-filled crystalline nylons are creep resistant; heat-deflection temperatures approach the crystalline melting points in dry resin. With good melt flow and rapid crystallization rates, nylons are easy to process. On the other hand, they exhibit high mold shrinkage, and dimensions and properties vary with the amount of absorbed water, which can be as high as 10% by weight; an increase in water content reduces the glass-transition temperatures. The lower the glass-transition temperature, the greater the creep under load. Mineral filling reduces but does not eliminate these problems. Thus part and mold designs have to compensate for dimensional changes and warpage.

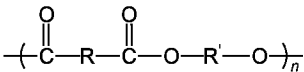
Table 5. Properties of Nylon Resin^a

Property	ASTM method	Nylon-6,6 Zytel 101 ^{b,c}	Nylon-6 Zytel 211 ^{b,c}	Trogamid T ^{d,e}
HDT at 1.82 N/mm ² , °C ^f	D648	194	129	124
melting point, °C	D2117	270	228	
notched Izod at 20°C, 3.2-mm thickness, J/m ^g	D256	53	80	65
flex modulus at 20°C, N/mm ^{2f}	D790	540–2800	1000	2700
specific gravity	D792	1.14	1.13	1.12
flammability rating, UL 94		V-2	HB	V-0

^a Ref. 58.
^b As molded; crystalline
^c Ref. 59.
^d Amorphous yellow resin; light transmission 85–90%.
^e Ref. 60.
^f To convert N/mm² to psi, multiply by 145.
^g To convert J/m to ftlbf/in., divide by 53.38.

Amorphous nylons are transparent. Heat-deflection temperatures are lower than those of filled crystalline nylon resins, and melt flow is stiffer; hence, they are more difficult to process. Mold shrinkage is lower and they absorb less water. Warpage is reduced and dimensional stability less of a problem than with crystalline products. Chemical and hydrolytic stability are excellent. Amorphous nylons can be made by using monomer combinations that result in highly asymmetric structures which crystallize with difficulty or by adding crystallization inhibitors to crystalline resins such as nylon-6 (61). Crystalline nylons are processed by injection molding and extrusion. The extrusion products are mostly films. Coated or laminated with moisture-barrier resin, eg, PVDC or PE polymers, they are used for meat wrappings and shaped food containers. For injection moldings, extruded pellets are used, available in neat, mineral-filled, glass-reinforced, pigmented, and impact-modified grades. Notched Izod impact resistance can be increased from the usual 100–150-J/m (1.9–2.8-ftlbf/in.) (under ambient moisture conditions) to the 1000-J/m (19-ftlbf/in.) range (62). Injection-molded parts are used for small mechanical, electrical, and building construction applications. In the automotive area, nylon resins are used for interior, exterior, and under-the-hood applications. Parts include filter bowls (amorphous nylon resins), window cranks, electrical connectors, fuse boxes, and speedometer gears. The electrical and electronic industries use various nylon resins for connectors, switches, and clamps. Mechanical applications include metal replacement in parts such as gears, sprockets, and wedges. The construction industry uses injection molded parts for fasteners, hardware, and power tools. For many applications, nylons compete with polyester and PC resins. Small volumes (about 5% of total nylon engineering applications) of nylon-11 [25035-04-5] and nylon-12 [24937-16-4] are used for solvent resistant parts such as automotive fuel line components. Over 565,000 t/yr of nonfiber crystalline nylons is sold worldwide (63). Since markets are controlled by the economy, a modest growth of 5–8%/yr is expected. Although currently only ca 900 t/yr of amorphous nylons is sold worldwide (64), a growth rate of 10% is expected because of increased research activity. Currently, the amorphous nylon resins compete with PEI and polyesters in many applications.

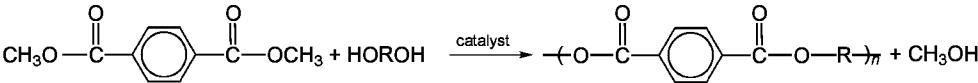
Polyester Resins. The general formula of engineering thermoplastic polyester resin may be given as follows:



Most polyesters (qv) are based on phthalates. They are referred to as aromatic-aliphatic or aromatic according to the copolymerized diol. Thus poly(ethylene terephthalate) [25038-59-9] (PET), poly(butylene terephthalate) [24968-12-5] (PBT), and related polymers are termed aromatic-aliphatic polyester resins, whereas poly(bisphenol A phthalate)s are called aromatic polyester resins or polyarylates; PET and PBT resins are the largest volume aromatic-aliphatic products. Other aromatic-aliphatic polyesters (65) include Eastman Kodak's Kodak resin, which is a PET resin modified with isophthalate and dimethylolcyclohexane. Polyarylate resins are lower volume specialty resins for high temperature (HDT) end uses (see HEAT RESISTANT POLYMERS).

Aromatic–Aliphatic Polyester Resins. Unlike most other classes of engineering plastics, which are made by only a few manufacturers, aromatic-aliphatic polyester resins are produced and compounded by several dozen firms (66). The aliphatic polyester resin marketplace is characterized by wide product differentiation and competition. Some firms make only a few hundred tons per year and presumably retain profitability because of the availability of low cost monomer and the simplicity of the processes employed. Low investment and low manufacturing costs are possible even for small-volume operations.

Low molecular weight PET and PBT resins are made by melt processes. For higher molecular weight resins, both melt processes or solid-state polymerization are used. Although terephthalic acid can be directly esterified, the most common process involves transesterification of dimethyl terephthalate with ethylene glycol or 1,4-butanediol in the presence of trace amounts of metal ion catalysts (67,68).



The methanol is collected overhead and the neat resin is extruded from the reactor in a batch or continuous process. The larger volume firms employ continuous operations with outputs per line of 11,000–14,000 t/yr (68).

The engineering applications of PET resins include blow-molded bottles, films, molding, and extrusion. Resins made for the latter two uses and related purposes are called molding resins in this article. The huge volumes of PET resin used for textile filaments and industrial fibers, eg, tire cord, are not included here. The PBT resins are mainly used for molding and related applications.

Bottle-grade PET resin has a $[\eta]$ range of 0.72–0.85 dL/g. Gels and particulates are minimized to achieve low intrinsic viscosity, and the manufacturing process is different than that for film or molding resins. Instead of continuing to heat and stir the low molecular weight resin to build a higher molecular weight, the resin is extruded as pellets at an $[\eta]$ of 0.6 dL/g. The pellets are heated in a hot nitrogen gas stream to promote further polymerization in the solid state. The temperatures used are lower than those required for molding, and shear effects are eliminated (66).

In 1977, consumption of PET resin in bottle applications was dramatically increased when the FDA banned competing acrylonitrile resins owing to toxicity considerations (recently rescinded) (69) and when the 2 L bottle was accepted for beverage sales worldwide (70). The carbon dioxide barrier properties of PET are sufficient to provide the six-month shelf life necessary for carbonated beverages (qv) (see also BARRIER POLYMERS).

Most PET bottles are produced by injection blow molding (71) the resin over a steel-core rod. The neck of the bottle is formed with the proper shape to receive closures and resin is provided around the temperature-conditioned rod for the blowing step. The rod with the resin is indexed to the mold, and the resin is blown away from the rod against the mold walls, where it cools to form the transparent bottle. The finished bottle is ejected and the rod is moved again to the injection-molding station. This process is favored for single cylindrical bottles, but cannot be used for more complex shapes such as bottles with handles.

The PET resins used for film and sheet applications are exceptionally clean resins with $[\eta]$ ranging from 0.6 to 1.0 dL/g. Most of the large volume general-purpose packaging film is resin of the lower (textile-fiber range) viscosity range. General-purpose as well as high tensile strength films are made by stretching the film longitudinally or laterally with timed periods of heating and cooling. Films stretched along two axes are termed biaxially oriented. The higher the extent of stretching, the higher the tensile strength and the lower the elongation to break. Stretching increases crystallinity, and heating anneals the stretched film and confers dimensional stability between 70 and 150°C. For heat sealing, the film is often coated with poly(vinylidene chloride) (PVDC) (72). Applications of biaxially oriented films include packaging films, heat-shrink film, magnetic-coated tapes, and computer floppy disks (72).

Both PET and PBT resins are used for moldable and extrudable engineering resins; intrinsic viscosity is typically in excess of 0.7 dL/g. Firms manufacturing molding resins either produce a certain molding-grade resin or use resins for compounding. This is especially true for PET resin products. The PBT resin was the first to be commercialized for injection molding in 1970, because PET crystallized slowly. The crystallinity of molded PET parts varied from molding to molding and changed over time, which often resulted in cracking. In 1978, Du Pont introduced the PET molding compounds in its Rynite line, using nucleating agents which enhanced the crystallization rates (73).

PET and PBT molding resins are available as neat resins or compounded with other resins (including each other), pigments, fillers, flame retardants, and reinforcements. Filled and reinforced resins are most commonly used for large parts and those with flat surfaces. Fillers retard mold shrinkage and the warpage characteristic of neat resin. However, fillers increase notched impact sensitivity and some grades contain additional impact modifiers. Another advantage of the filled resins is the increase in heat-deflection temperature from the glass-transition temperature to values approaching the melting point. The properties given in Table 6 are for opaque cream-colored resins with good-to-excellent solvent resistance and good hydrolytic stability.

Table 6. Properties of Aromatic–Aliphatic Polyester Resins^a

Property	ASTM method	Valox 325 ^{b,c}	Valox DR-51 ^{c,d}	Rynite 530 ^{e,f}
HDT at 1.82 N/cm ^{2g} , °C	D648	130	160	224
notched Izod at 20°C, 3.2-mm thickness, J/m ^h	D256	50	55	90
flex modulus at 20°C, N/mm ^{2g}	D790	2300	4530	6050
specific gravity	D792	1.31	1.41	1.56
flammability rating, UL 94			V-0	HB

^a Ref. 74.
^b Neat PBT resin.
^c Ref. 75.
^d PBT resin, 15% glass-reinforced.
^e PET resin, 30% glass-reinforced.
^f Ref. 76.
^g To convert N/mm² to psi, multiply by 145.
^h To convert J/m to ft-lbf/in., divide by 53.38.

PET and PBT resins often compete for the same applications. The PET resins are considered in cases where higher heat-deflection temperatures and greater rigidity are desired. The PBT resins have the advantages of faster crystallization and reduced moisture absorption, which allows faster molding cycles and reduced drying time. Typical applications include automotive parts such as distributor caps, air-blower deflectors, connectors and housings, and exterior parts, eg, bowls and fender extensions. The solvent resistance and high heat-deflection temperatures of PBT and PET resins permit replacement of automotive metal parts by those made from these resins. Blends of PBT with PC resins are used for auto bumpers. Highly ductile, specially formulated polyester resins will likely compete for other large exterior parts, including doors, fenders, and quarter panels. Other markets are electrical and electronic parts (which use flame-retarded UL 94 V-0 grade), home appliances, furniture, plumbing valves, pump housings, brackets, medical components, and parts for sports equipment.

Processing is similar to other engineering plastic resins. Drying is necessary before extrusion or molding. Special drying precautions are required for PET products to prevent degradation and splay.

Both PET and PBT resin serve some of the same markets as other plastics. Nylon, unsaturated polyester fiber glass, phenolic, PC, and polyarylate resins sometimes compete for the same molded part.

Bottle-grade PET resin with marginal gas barrier properties competes with PC resins which can be coextrusion blow-molded with specialty barrier resins. PET films afford the higher heat-deflection temperature and dimensional stability required by small electronic circuit constructions. Polyolefin films compete for the same markets and are less expensive. Nonetheless, combined PET and PBT polyester resins are sold at a rate of 800,000 t/yr worldwide (Table 7) and are expected to have an average annual volume growth rate of 7–12% (66,77); in Europe, bottle-grade PET is growing at a rate of 23% (66). Growth rates for molding resins could increase if usage for large exterior automotive parts increases (78).

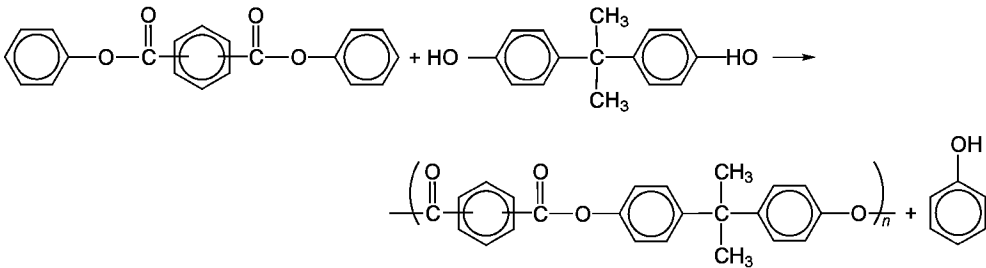
Table 7. Terephthalate Resin Sales, 10⁵ t/yr

Area	PET		PET and PBT molding
	Bottle	Film	
Western hemisphere	2.55	2.15	0.55
Europe	0.50	0.94	

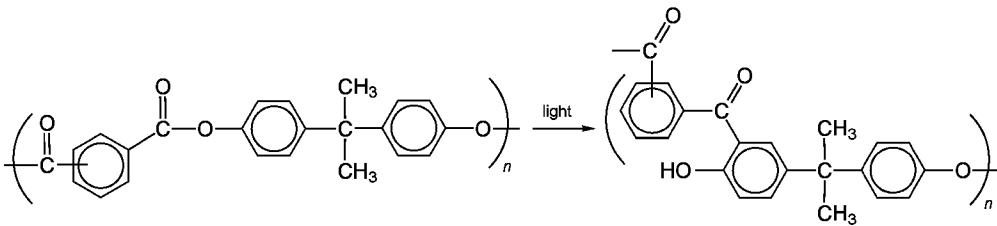
Asia	0.27	0.93	0.64
Total	3.32	4.02	1.19

Aromatic Polyester Resins. A number of attempts to commercialize aromatic polyesters have been made. However, for reasons such as moldability, clarity, and color stability, significant market share has never been achieved.

Commercial aromatic polyester resins or polyarylates are a combination of bisphenol A with isophthalic acid or terephthalic acid (79). The resins are made commercially by solution polymerization or melt transesterification (47).



Blends of iso- and terephthalates give amorphous, transparent resins, mostly yellow in color. Heat-deflection temperatures are higher than those of 100% PC resins and depend on the iso- to terephthalate ratio. For example, a resin with a 1:1 ratio has a value of 160°C. These resins are flame retardant; mechanical and electrical properties are similar to those of PC resins. The notched Izod impacts are lower at 150–300 J m (4.7–5.6 ftlb/in.), even in thick sections. Long-term uv radiation stabilities are excellent, but yellowness increases during initial exposure owing to photo-Fries rearrangements (80), wherein *o*-hydroxy-benzophenone units are produced along the polymer chains.



The chromophores are yellow and act as uv screening agents. The reactions occur at surfaces and uv penetration is avoided. Mechanical properties are thereby protected (47).

Polyarylates are sensitive to heat. Although mechanical properties are not much affected, colors darken. Properties are given in Table 8. Hydrolytic stability and resistance to organic solvents are fair.

Table 8. Properties of Aromatic Polyester Resins^a

Property	ASTM method	Ardel D-100	Durel 400
HDT at 1.82 N mm ^{2b} , °C	D648	174	170
notched Izod at 20°C, 3.2-mm thickness, J m ^c	D256	210	275
flex modulus at 207°C, N mm ^{2b}	D790	2100	2300
specific gravity	D792	1.21	1.21
flammability rating, UL 94		V-0	V-0
color		yellow	colorless
light transmission T, %		75	84–88

^a Refs. 47,81.

^b To convert N mm² to psi, multiply by 145.

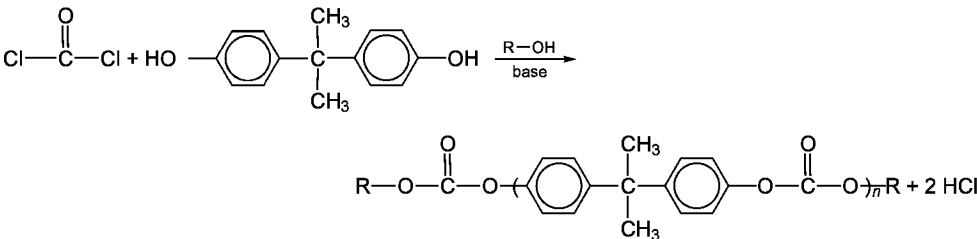
^c To convert J m to ftlb/in., divide by 53.38.

Polyarylates have been employed for electrical and electronic components, firefighter helmets, and applications requiring higher heat-deflection temperatures than PC resins. Polyarylates were first used for lighting, especially for small automotive lenses and sodium light outdoor lamps. However, the inherent yellow color and heat-induced darkening have limited applications.

Although several companies have produced polyarylates, Unitika in Japan and Hoechst-Celanese are the only ones active in the marketplace. Unitika resin is marketed in the United States by Amoco Performance Polymers under the Ardel trademark. The reason for this decline may be the competition of PC, polyester carbonates, and polysulfone resins. Worldwide sales are quite small, probably less than 1000 t yr.

Polycarbonate (PC) Resins. Polycarbonates (qv) based on bisphenol A are sold in large quantities. Other bisphenols can be incorporated, but do not give the same favorable combination of properties and cost (82). Small quantities of PC based on tetramethylbisphenol A are used as blending resins (83) and polyester carbonate copolymers are used for applications requiring heat-deflection temperatures above those of standard PC resins (47).

Bisphenol A Polycarbonate Resins. These resins are manufactured by interfacial polymerization (84,85). A small amount of resin is produced by melt-polymerization of bisphenol with diphenyl carbonate in Russia and the People's Republic of China. Melt technology continues to be developmental in Japan and the West, but no commercial activities have started-up to date, although some were active in the late 1960s. No reports of solvent-based PC manufacture have been received.



The terminal R groups can be aromatic or aliphatic. Typically, they are derivatives of monohydric phenolic compounds including phenol and alkylated phenols, eg, *t*-butylphenol. In interfacial polymerization, bisphenol A and a monofunctional terminator are dissolved in aqueous caustic. Methylene chloride containing a phase-transfer catalyst is added. The two-phase system is stirred and phosgene is added. The bisphenol A salt reacts with the phosgene at the interface of the two solutions and the polymer "grows" into the methylene chloride. The sodium chloride by-product enters the aqueous phase. Chain length is controlled by the amount of monohydric terminator. The methylene chloride–polymer solution is separated from the aqueous brine-laden by-products. The facile separation of a pure polymer solution is the key to the interfacial process. The methylene chloride solvent is removed, and the polymer is isolated in the form of pellets, powder, or slurries.

Although some 12 firms produce PC resin worldwide, Bayer AG along with its subsidiary Miles, Inc., General Electric Company, and The Dow Chemical Company are the principal producers.

PC resins are glassy, amorphous polymers with little color and excellent optical properties. The high molecular weight resins are tough, with notched Izod impact values of 600–800 J m (11.2–15.0 ftlb/in.). Heat-deflection temperatures of unfilled grades are near 130°C; glass reinforcement increases these values by 10°C. Flex moduli above 2400 N mm² (348,000 psi) allow PC resins to be formed into large rigid parts. Below 100°C, creep under load is low, as is predictable from the high glass-transition temperatures. PC resin products have good hydrolytic stability but should not be used for applications requiring repeated autoclaving. Like most amorphous resins, they are susceptible to attack by some organic solvents. Although PC resins have good resistance to fruit juices, aliphatic hydrocarbons, aqueous alcohol, and strong battery acids, they dissolve in chlorocarbon solvents and stress-crack in contact with ketonic solvents such as acetone and methyl ethyl ketone. Properties are given in Table 9.

Table 9. Properties of Polycarbonate (PC) Resins^a

Property	ASTM method	Lexan 101 ^b	Medlon 8320 ^{c,d}	Lexan 4701 ^{b,e}
HDT at 1.82 N mm ^{2f} , °C	D648	132	142	163
notched Izod at 20°C, 3.2-mm thickness, J m ^g	D256	600	100	500
flex modulus at 20°C, N mm ^{2f}	D790	2300	3380	2300
specific gravity	D792	1.21	1.27	1.2
flammability rating, UL 94		V-2	V-1	HB
light transmission T, %		85	opaque	85

^a Ref. 86.
^b Ref. 87.
^c 20% glass-reinforced.
^d Ref. 88.
^e Polyester cabonate.
^f To convert N mm² to psi, multiply by 145.
^g To convert J m to ftlb/in., divide by 53.38.

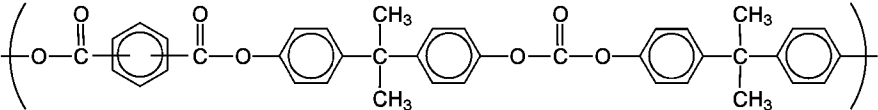
PC resin materials are sold mainly as injection-moldable pellets available in transparent, flame-retarded, mineral-filled, reinforced, and impact-modified grades. PC resins are also used for foam molding. Some grades contain blowing agents activated by heat; others receive the blowing agents by online gas or volatile liquid injection during foam molding. Sheet products are available in forms ranging from uv-stabilized construction grades to multilayer laminates designed for bullet-resistant glazing and lightweight twin-wall extrusions. Both blown and extruded film are produced. Some branched PC resins are made specifically for blow-molding applications by introducing polyfunctional monomers such as trimellityl trichloride during polymerization; branching increases melt strength.

Processing PC resins by extrusion or injection-molding methods requires melt temperatures of 290–320°C. High melt viscosity at low shear rates prevents mold flash and drool. At injection shear rates, apparent viscosities decrease, and easy melt flow allows manufacture of large, complex parts.

PC resins find applications in many areas. Sheet products are used for glazing and thermoformed products such as signs, cases, and furniture. PC resin film is suitable for silk-screen printing and finds uses in illuminated consoles and soft-touch controls. PC resins are utilized for packaging as components of multilayer extruded sheet and film and in coextruded blow-molded containers. Injection molding consumes 80% of production. Applications include electrical and electronic parts such as housings, connectors, and fuses, appliance parts, power-tool housings, parts for the transportation industries ranging from automobile dashboards to headlamps. Sports equipment requiring high impact and other uses might be categorized under metal or glass replacement. PC blow-molding resins are used for containers, including large water bottles. Blending trades off on toughness and impact resistance with properties of other resins such as solvent resistance.

Current worldwide PC resin sales are ca 290,000 t/yr (70,89); annual growth rates are expected to be near 10% (90).

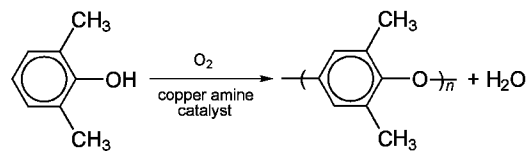
Polyester Carbonate Copolymers. Polyester carbonate resins have molecular structures composed of iso- and terephthalate units in conjunction with the standard bisphenol A PC moieties.



A convenient method of compositional designation uses molar percentages of ester and carbonate linkages coupled with the molar percentages of iso- and terephthalate units in the polymer. A 70% ester, 30% carbonate polyester carbonate with 60 parts of isophthalate and 40 parts of terephthalate is designated 70(60/40)30. Similarly, a standard PC resin is 0(0/0)100, and a polyarylate resin composed of a 1:1 molar ratio of iso- to terephthalate units esterfied with bisphenol A is designated 100(50/50)0.

Polyester carbonate resins are made by the interfacial process described for standard PC resins. The phthalate units are introduced by reaction of the appropriate phthaloyl dichlorides concurrent with the phosgenation. At present, Bayer, GE, and Miles produce polyester carbonate resins (47); sales volume is low, probably ca 100 t/yr. Polyester carbonates are used primarily in applications requiring 5–25°C higher heat-deflection temperature and better hydrolytic performance than are provided by standard PC resins. Properties are given in Table 9.

Poly(phenylene ether) Alloys. Poly(phenylene ether) resins (91), composed of phenolic monomers, have a very high *T_g*. The commercial resins are based on 2,6-dimethylphenol. The resin is produced by oxidative polymerization in toluene solution over an amine catalyst (see also POLYETHERS, AROMATIC).



With the glass-transition temperatures above 260°C, neat poly(phenylene ether) resins are difficult to process; melting does not begin below 255°C. At extrusion or injection-molding temperatures of 300–350°C, melt flow is stiff and special precautions must be taken to minimize oxidation. Poly(phenylene ether) resins are soluble in all proportions with polystyrene resin and other styrenic polymers, which allows blends or alloys (see blends and alloys below) to be made. The higher the PS content, the lower the heat-deflection temperature and the easier is alloy processing. Without further modifications, however, these alloys are brittle. Rubber-impact modifiers improve toughness. HDT and cost are controlled by blending ratios of PPO and PS. In practice, the poly(phenylene ether) resin–PS resin alloys are compounded with many ingredients, eg, stabilizers and flame retardants. Thus, although the alloys are clear and amorphous, the compounded alloys are opaque.

Typical poly(phenylene ether) resin–styrene resins are cream colored and are characterized by excellent melt flow. They are easily foamed either by an internal heat-activated blowing agent or by in-line injection of a volatile material. Flame retardance is conferred by halogen-free compounds such as aryl phosphates. Electrical properties are outstanding and moisture absorbance is low. With excellent dimensional stabilities, the alloys can be made into large parts. Although hydrolytic stability is good, organic solvent resistance is poor. This has limited applications, such as large automotive parts. Final finishing must take chemical resistance into consideration. Another problem area is color stability to uv radiation. Properties are given in Table 10.

Table 10. Properties of Poly(phenylene ether)-based Resins^a

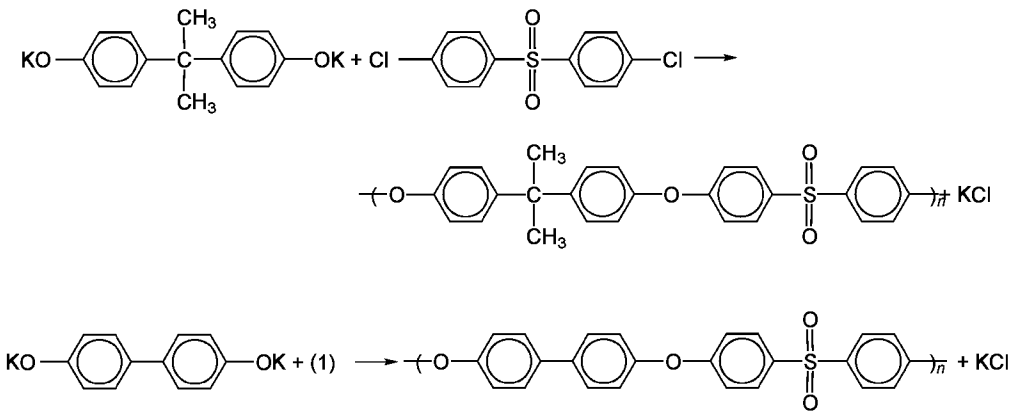
Property	ASTM method	Noryl Se100 ^b	Prevex VQA ^c	Noryl GFN3 ^{b,d}
HDT at 1.82 N mm ^{2e} , °C	D698	100	129	135
notched Izod at 20°C, 3.2-mm thickness, J m ^f	D256	250	268	120
flex modulus at 20°C, N mm ^{2e}	D790	2500	2600	7700
specific gravity	D792	1.10	1.06	1.36
flammability rating, UL 94		SE-1	V-1	SE-0

^a Ref. 92.
^b Ref. 93.
^c Ref. 94.
^d Glass-reinforced.
^e To convert N mm² to psi, multiply by 145.
^f To convert J m to ftlb/in., divide by 53.38.

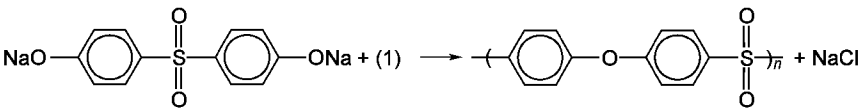
Resins other than styrene are being investigated for blending. For example, poly(phenylene ether) resins blended with polyamide resins give dimensionally stable, solvent-resistant materials (95). This illustrates the use of compatibilization to overcome insolubility. The materials are promising and could have an impact on large-part injection-molding and blow-molding technologies. They are currently used in automotive vertical panels. Pellets for injection-molding, extrusion, and blow-molding applications include mineral-filled, glass-filled, pigmented, and flame-retarded grades. The alloys can be processed into parts requiring hydrolytic and autoclave stability, such as desalinization filter frames and medical devices. Foamed products are used in cabinets for business machines, computers, and copiers. Automotive seat backs are made by blow molding and replacements for metal parts by injection molding. Worldwide sales of poly(phenylene ether)–styrene resin alloys are 100,000–160,000 t/yr (47,96); annual growth rates are ca 9% . Other resin, particularly acrylonitrile–butadiene–styrene (ABS) polymers and blends of these resins with PC resins, compete for similar applications.

High Performance Engineering Resins

Polysulfone Resins. Commercially important polysulfones are aromatic, ie, in the generalized formula for the repeating unit R and R' both contain aromatic rings (see POLYMERS CONTAINING SULFUR, POLYSULFONE RESINS). They all possess ether linkages as well, so that use of the designations polysulfone, polyarylsulfone (PAS), and polyethersulfone (PES) is somewhat arbitrary. Union Carbide introduced the first commercial polyarylsulfone resin, Udel, in 1966 and later introduced Radel polyarylsulfone. The business and patents rights were acquired by Amoco Corp. in 1986. Other polysulfone products have been marketed by BASF and ICI (Victrex PES); the latter announced in 1991 it was withdrawing from the business. The resins are made by batch processes employing Friedel-Crafts reactions or nucleophilic aromatic substitution. Udel resin and Radel R resin are produced by the nucleophilic displacement of chloride on 4,4'-dichlorodiphenyl sulfone by the potassium salts of bisphenol A and 4,4'-biphenol, respectively (97):



Radel A 400 polyarylsulfone resin is made from a proprietary blend of bisphenols (believed to include hydroquinone) and dichlorodiphenyl sulfone. Victrex resin is made by other displacement reactions.



Udel is a slightly yellow but transparent engineering thermoplastic. It has low flammability and smoke emission and good electrical properties. It has excellent resistance to water, steam, and alkaline solutions. Specific uses for Udel include microwave cookware, beverage dispensers, coffee brewers, cookware, hair dryers, corn poppers, and steam table trays. Its steam resistance makes it particularly fit for a dishwasher environment. Properties of polysulfone resins are given in Table 11.

Table 11. Properties of Polysulfone Resins

Property	PES Victrex 200P	PAS Radel A-400	PSO Udel
specific gravity	1.37	1.37	1.24
tensile strength, MPa ^a	82.8	82.8	70.3
elongation at break, %	40–80	40	50–80
flexural modulus, GPa ^a	2.55	2.69	2.69
notched Izod, J m ^b	85.4	85.4	69.4
heat deflection temperature, °C at 1.82 MPa ^a	203	204	174
flammability rating, UL 94	V-0	V-0	V-2

^a To convert MPa to psi, multiply by 145.
^b To convert J m to ftlbf in., divide by 53.38.

Blends of the polysulfone resin have been made with ABS, poly(ethylene terephthalate), polytetrafluoroethylene (PTFE), and polycarbonate. These are sold by Amoco under the Mindel trademark. Additional materials are compounded with mineral filler, glass, or carbon fiber to improve properties and lower price.

Amoco Performance Products (Atlanta, Ga.), is the sole supplier of polysulfone, Udel, polyarylsulfone, Radel A, and polyphenylsulfone, Radel R. ICI Advanced Materials (Exton, Pa.), is the sole domestic supplier of the polyethersulfone, Victrex PES, but announced in 1991 it was withdrawing from the business.

In 1989 BASF announced its intention to market in the United States a polysulfone, Ultrason S, and a polyethersulfone, Ultrason E. Both materials would likely be made in Germany. World consumption and prices of polysulfones are given in Table 12.

Table 12. World Consumption of Polysulfones

Resin	Consumption, t	Price, \$ kg
polysulfone	8200	8.60 ^a
polyethersulfone	3600	9.70 ^b
polyarylsulfone	400	
polysulfone blends		5.70–6.60 ^c

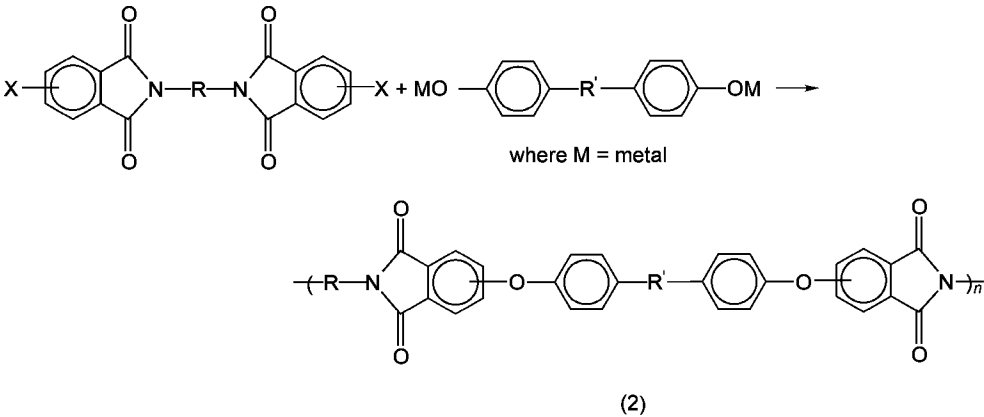
^a Udel, neat.
^b Victrex PES, neat.
^c Mindel, compounded.

Medical uses for Udel resin include surgical trays, nebulizers, flow controllers for blood, and respiration regulators. Transportation applications center around automotive fuse housings, electrical connectors, and switches. Electrical and electronic end uses include coil bobbins, housings, connectors, bushings, capacitor film, and business machine parts. Finally, water, heater dip tubes, milking machine parts, pollution control equipment, and some filtration membranes are made.

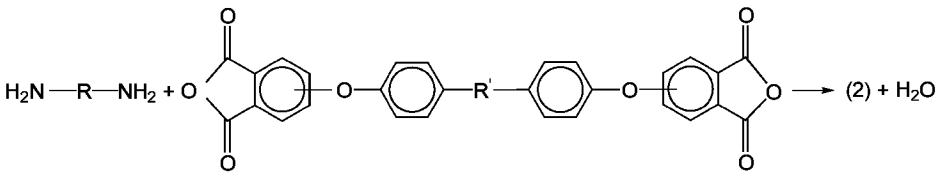
Victrex (PES) and Radel (PAS) serve the same end uses and are, like Udel, transparent, slightly yellow polymers exhibiting heat deflection temperatures about 28°C higher than Udel PSO. Their main advantage in electrical and electronic end use is high heat resistance. They also have excellent water and chemical resistance; like all polysulfones they are steam sterilizable. Electrical and electronic end uses represent the largest market segment. Applications include connectors, switches, bobbins, and sleeves.

In automotive and aerospace end uses, the applications are also often electrical. Polysulfones do not have as good solvent resistance as poly(phenylene sulfide). They perform well in hydrocarbons like gasoline and oil, or in antifreeze, but are attacked by the alcohol-blend fuel mixtures. This may limit their under-the-hood applications.

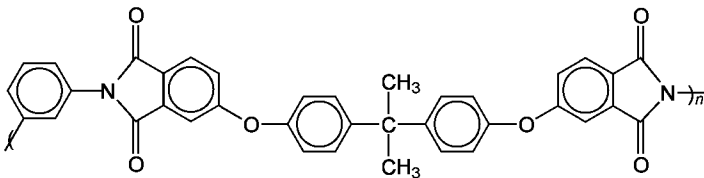
Polyetherimide Resins. Polyetherimide resins (PEI) (7) were commercialized during the 1980s (see POLYIMIDES). They are produced by an unusual nucleophilic substitution process:



or by a conventional condensation of diamines and dianhydrides:



Many different resins can be made by varying R and R', as in Ultem resins.



Ultem 1000 resin

Ultem PEI resins are amber and amorphous, with heat-distortion temperatures similar to polyethersulfone resins. Ultem resins exhibit high modulus and are stiff yet ductile. Light transmission is low. In spite of the high use temperature, they are processible by injection molding, structural foam molding, or extrusion techniques at moderate pressures between 340 and 425°C. They are inherently flame retardant and generate little smoke; dimensional stabilities are excellent. Large flat parts such as circuit boards or hard disks for computers can be injection-molded to maintain critical dimensions.

PEI polymers exhibit minimal creep under load. For example, unreinforced Ultem 1000 maintains part dimensions over thousands of hours at room temperature at a loading of 34 N mm² (4930 psi). PEI resins are available with HDTs ranging to 220°C (Ultem 5000). PEI copolymers with silicone rubber allow for flame-retardant, high temperature applications such as plenum wire coatings (Siltem). Reinforcement of PEI resins improves their temperature performance. Table 13 compares unreinforced and 20% glass-reinforced Ultem resins.

Table 13. Properties of Polyetherimide (PEI) Resin^a

Property	ASTM method	Ultem 6200 ^b	Ultem 1000 ^c
HDT at 1.82 N mm ^{2d} , °C	D648	223	200
notched Izod at 20°C, 3.2-mm thickness, J m ^e	D256	80	50
flex modulus at 20°C, N mm ^{2d}	D790	6550	3240
specific gravity	D792	1.43	1.7
flammability rating, UL 94		V-0	V-0

^a Refs. 98,99.

^b 20% glass-reinforced.

^c Neat.

^d To convert N mm² to psi, multiply by 145.

^e To convert J m to ftlbf in., divide by 53.38.

The Ultem PEI resins compete with PAI, polyarylethersulfone, nylon, and polyester resins in certain markets. General Electric Co. is the sole U.S. manufacturer of PEI resins. High cost coupled with stiff competition from metals and ceramics have limited growth.

Poly(phenylene sulfide) Resins. Poly(phenylene sulfide) (PPS) resin is manufactured from *p*-dichlorobenzene and sodium sulfide in a dipolar aprotic solvent (100).



The resin is recovered by precipitation as white powder, which is then purified by extraction with 2-propanol, dried, and then generally compounded with fiber glass and mineral filler. In this Phillips process the white powder undergoes "curing" when heated. Manufacturers heat the resin at 175–280°C in air for 2–4 h to form a chain extended, branched to lightly cross-linked product. If the resin powder is heated above its melting point at 285°C in air for some time, it darkens, gels, and ultimately forms a black thermoset solid. This curing process is useful for thin coatings, 25–50 μm. Since properties of PPS depend greatly on crystallinity, it is critical that this be controlled during manufacturing by controlling the thermal history in order to obtain the desired characteristics in the finished product. For instance, low mold temperatures lead to products with lower crystallinity and HDT. An alternative process (Fortron Industries) can produce a linear PPS product through direct polymerization without a curing step.

The polymer (PPS) possessing high crystallinity exhibits greater resistance to creep than the amorphous material. On the other hand, the latter has higher elongation. Most poly(phenylene sulfide) resin is sold as a compound of glass or mineral filler. The main advantage of PPS lies in its superior heat deflection and continuous service temperatures. PPS properties are listed in Table 14. PPS is an excellent electrical insulator and is inherently flame resistant. Its chemical resistance is outstanding even at elevated temperatures, except in oxidizing agents, strong acids, and chlorinated hydrocarbons (see POLYMERS CONTAINING SULFUR, POLY(PHENYLENE SULFIDE)).

Table 14. Properties of Poly(phenylene sulfide) (PPS) Resin^a

Property	ASTM method	PPS resin ^b	Ryton R-4 ^{c,d}
HDT at 1.82 N mm ^{2e} , °C	D698	111	241
notched Izod at 20°C, 3.2-mm thickness, J m ^f	D656	11	63
flex modulus at 20°C, N mm ^{2e}	D790	3,845	11,000
specific gravity	D792	1.36	1.6
flammability rating, UL 94		V-0	V-0

^a Annealed; properties are influenced by the degree of crystallinity (101).

^b Neat.
^c 40% glass-reinforced.
^d Ref. 102.
^e To convert N/mm² to psi, multiply by 145.
^f To convert J/m to ftlb/in., divide by 53.38.

PPS resins are chiefly used for injection molding. The melt flow of the glass-filled resins is very stiff, and high injection pressures are required. Mold surface wear is heavier than for most other engineering plastics. Molding melt temperatures are near 330°C; for optimum surface gloss and impact strength, mold temperatures of 130°C should be used. The resins are brown to brown-black.

Injection-molding pellets are available in a number of grades including neat, pigmented, mineral-filled, and glass-reinforced. Applications include chemical-resistant parts for pumps and processing equipment such as impellers. The high heat resistance of glass-reinforced grades allows application in large exterior light reflectors and automotive engine components.

PPS resins must compete with PEI and phenolics. There are two domestic manufacturers of poly(phenylene sulfide): Phillips and Fortron Industries. Worldwide there is currently large overcapacity (Table 15). Four Japanese companies, ie, Toso Susteel, a joint venture of Toso/Hodogaya Chemical; Toray; Toprene, a joint venture of Toto Kasei/Toren Petrochemical; and Kureha Chemical have a combined capacity of 82,500 t. U.S. agents sell their materials in the U.S. markets: General Electric sells for Toso Susteel; Soltex Polymer, part of Solvay, Belgium, sells for Toprene; Hoechst-Celanese sells for Kureha. Prices for PPS resins and compounds range from \$8.80/kg for unreinforced resin to \$3.30/kg for 65% filled resins.

Table 15. World Production Capacity of Neat PPS Resin in 1990

Manufacturer	Capacity, t	Plant location
Toso Susteel/GE	3,000	Japan
Toray/Phillips	7,500	Japan
Phillips	7,500	United States
Toprene/Solvay	3,600	Japan
Kureha/Hoechst-Celanese	3,000	Japan
Hoechst-Celanese	3,600 ^a	United States
Bayer (Miles)	4,000	Germany
Total	31,600	

^a 1994.

Electrical and electronics are the largest end uses for PPS. Principal advantages of PPS are strength at high temperature, dimensional stability, flame and arc resistance, and precision moldability. Products are connectors, coil bobbins, relays, and switches. The growth of electrical and electronic end uses is 10%^o. Appliances are the second largest end uses for PPS resins as a replacement for phenolic, but also for metal parts.

Transportation end uses are expected to become a significant outlet. Products under development include an engine valve cover, as are various housings such as those for oil pumps, water pumps, starter motors, and certain transmission parts. These end uses employ PPS because it resists high temperatures and is also chemically resistant. Fuel system parts can employ the excellent chemical resistance of PPS, replacing nylon, if alcohol-based fuels are adopted to reduce emissions.

Fiber from PPS resins has been made in two forms. Monofilament is used in paper machine drier felts to replace polyester, which is attacked by the hot, corrosive conditions of papermaking. Staple fibers are made into filter bags for flue treatment, and are considered a growth area.

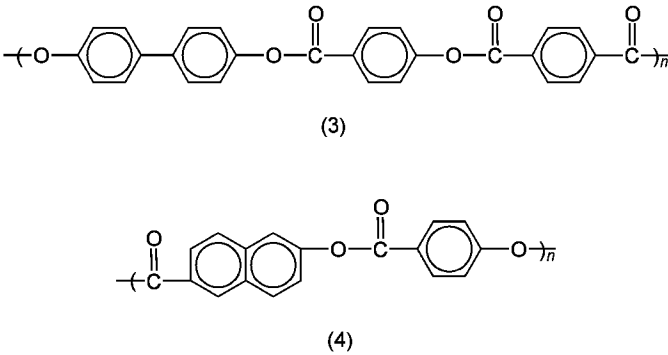
Chemical end uses employ the most exceptional property of PPS, chemical inertness. PPS is almost as chemically resistant as Teflon. It is used in pump impellers and housings and for down-oil-well end uses. New extrusion grades have been developed for potential piping end uses. Coatings of PPS are used extensively.

Liquid Crystalline Polymers. Liquid crystalline molding resins are, by and large, special aromatic resins possessing rigid linear molecular structures. These rigid rods form ordered crystal-like areas in the liquid state that coexist with amorphous areas. This molecular order can be improved by such processes as orienting a film or stretching a fiber as the polymer rods are then dragged against each other and are forced into a parallel array and achieve the close packing needed to "lock." Order is preserved by cooling, thus locking the structure in the solid state. LCP's rod-like molecular structure provides the strength and stiffness normally associated with fiber-reinforced thermoplastics. At temperatures of 200°C and above, LCPs maintain stiffness comparable with that of some engineering plastics at room temperature. These polymers are intended as metal replacement in electronic, automotive, and industrial markets.

In the ordered smectic or nematic phase, the rigid rods are arranged in parallel arrays that allow for close packing. The nematic phase is the most common type found with synthetic polymer molecules. The molecules' long axes are parallel, but there is no layering. Aromatic polymer chains that have stiff ester or amide linkages are ideal.

Four companies (trade name in parentheses), Amoco (Xydar), Hoechst-Celanese (Vectra), Du Pont, and Granmont (Granlar), make thermotropic LCPs for various types of extrusion and molding processes. Six companies have discontinued LCP materials that were either commercial or under development. These companies include ICI, BASF, Eastman, Bayer, General Electric, and Monsanto.

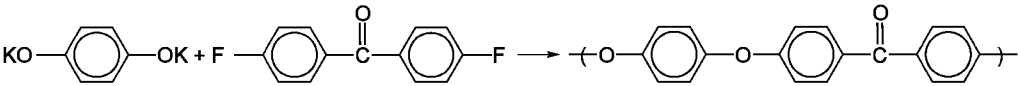
Chemically all the currently commercial materials are linear aromatic esters. Xydar (3) is a terpolymer of *p,p'*-biphenol, *p*-hydroxybenzoic acid, and terephthalic acid. Vectra (4) is a copolymer of *p*-hydroxybenzoic acid and 6-hydroxy-2-naphthoic acid.



Many different combinations of carboxylic acid and hydroxyl groups have been tested to form LCPs. An aromatic structure (benzene, naphthalene, anthracene, etc) is required that has its functional groups symmetrically arranged on opposite sides of the molecule. Examples are a 1,4-substituted benzene compound or 2,6-substituted naphthalene compound. These monomers are often complex and expensive molecules and account for a significant portion

of the high cost of the polymers. Markets have developed more slowly than initially anticipated in electronic, automotive, and industrial areas. The thermotropic LCP market was estimated at 1300 t in 1990 with a growth rate at 14% yr. Prices, once \$22 / kg, have fallen to \$17–20 / kg.

Polyetheretherketone Resin (PEEK). The resin was commercialized as Victrex PEEK by Imperial Chemical Industries, Ltd. (ICI) in the late 1970s and by Amoco Chemicals Corp. in the middle 1980s under the trade name Kadel. It is produced by both companies in the United States. Kadel is believed made by the displacement reaction of 4,4'-difluorodiphenyl ketone by the potassium salt of hydroquinone:



Other companies offering similar materials are Du Pont, BASF, and Hoechst-Celanese. PEEK is a difficult resin to manufacture because of batch procedures and insolubility.

The PEEK resin is gray, crystalline, and has excellent chemical resistance; *T_g* is ca 185°C, and it melts at 288°C. The unfilled resin has an HDT of 165°C, which can be increased to near its melting point by incorporating glass filler. The resin is thermally stable, and maintains ductility for over one week after being heated to 320°C; it can be kept for years at 200°C. Hydrolytic stability is excellent. The resin is flame retardant, has low smoke emission, and can be processed at 340–400°C. Crystallinity is a function of mold temperature and can reach 30–35% at mold temperatures of 160°C. Recycled material can be safely processed. Properties are given in Table 16.

Table 16. Properties of PEEK Resins^a

Property	ASTM method	Victrex PEEK resin ^b	Reinforced Victrex PEEK ^c	
			30% glass	30% carbon
HDT at 1.82 N/mm ² , °C	D648	165	282	282
notched Izod at 20°C, 3.2-mm thickness, J/m	D256	150	110	70
flex modulus at 20°C, N/mm ²	D790	2000	7700	1540
specific gravity	D792	1.32	1.44	1.32
flammability rating, UL 94		V-0	V-0	V-0

^a Ref. 103.

^c Ref. 105.

^b Ref. 104.

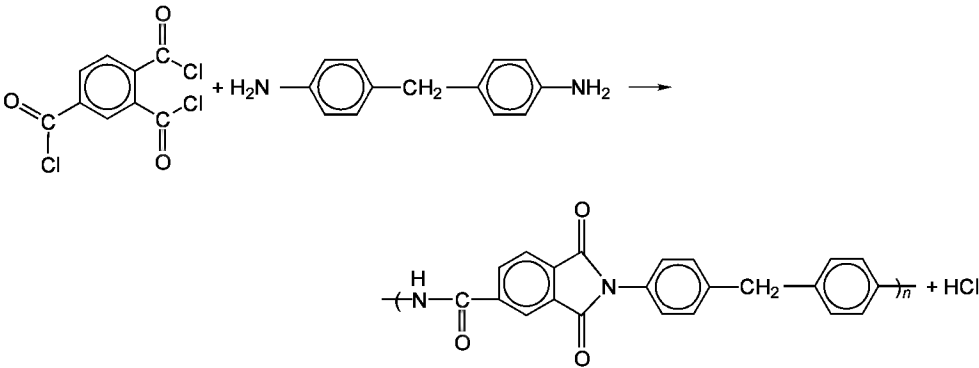
^d To convert N/mm² to psi, multiply by 145.

^e To convert J/m to ft-lbf/in., divide by 53.38.

The PEEK resin is marketed as neat or filled pellets for injection molding, as powder for coatings, or as preimpregnated fiber sheet and tapes. Applications include parts that are exposed to high temperature, radiation, or aggressive chemical environments. Aerospace and military uses are prominent. At present, polyamideimide (PAI) resin and poly(arylene sulfides) are the main competitors for applications requiring service temperatures of 280°C. At lower temperatures, polyethersulfones, amorphous nylons, and polyetherimides (PEI) can be considered.

Production of PEEK resin is less than 500 t/yr worldwide (106). Growth is expected to exceed 10% yr. Because of chemical and manufacturing constraints, the price, ca \$55/kg for neat resin, is likely to remain high.

Polyamideimide Resin (PAI). Amoco introduced the first commercial polyamideimide (PAI) resin, Tordon, in the early 1970s (107,108). It is produced by the solution condensation of trimellitic trichloride with methylenedianiline (109):



Tordon resin is an amber-to-gray amorphous polymer with an HDT of 274°C, flexural modulus of 4600 MN/m² (667,000 psi), and good impact resistance. Although it contains amide linkages, it is dimensionally much more stable under various humidity conditions than the standard crystalline nylons because of its elevated glass-transition temperature. Also, like PPS, Tordon appears to undergo "curing" with additional heat treatment. Like aromatic polyamides and imides, Tordon is flame retardant and has low smoke emission. Creep under load is low. Electrical, chemical, hydrolytic, and radiation stability are excellent. Because of these properties, PAI replaces metal in gears, connectors, and other mechanical parts. It is also used in electrical applications where high temperature and chemical resistance are required. Properties are given in Table 17.

Table 17. Properties of Tordon Polyamideimide (PAI) Resin^a

Property	ASTM method	Tordon 4203L ^b	Tordon 7130 ^{c,d}
HDT at 1.82 N/mm ² , °C	D648	274	278
notched Izod at 20°C, 3.2-mm thickness, J/m	D256	135	45
flex modulus at 20°C, N/mm ²	D790	5,110	20,000
specific gravity	D792	1.38	1.42
flammability rating, UL 94		V-0	V-0

^a Ref. 110.

^b Ref. 111.

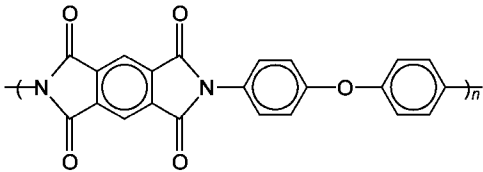
^c 30% graphite-reinforced.
^d Ref. 112.
^e To convert N mm² to psi, multiply by 145.
^f To convert J m to ftlb/in., divide by 53.38.

The PAI resin is utilized for small, simple parts. Because of its stiff melt flow, the resin is available in compression moldable powder. Grades include neat, filled, and glass-reinforced resins; it is also available in solution for use as an enamel. Because of high prices (\$33–44/kg) the market is limited, estimated at 150–300 tons for the molding resin. Resins competing with PAI include PEI, polyethersulfones, and thermoset resins.

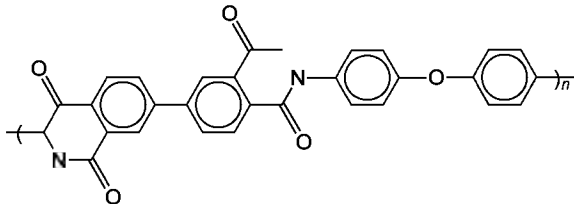
Other Polyimide Resins. The polyimide category, which includes both the polyetherimide and the polyamide resins discussed earlier, can be broken down into three subdivisions.

Nonthermoprocessible Condensation Polyimides. These are obtained from condensation of aromatic dianhydrides with aromatic diamines. They are linear noncross-linked resins but their rigid chain structure and strongly hydrogen-bonded character leads to systems which do not melt or soften before decomposition.

Kapton resin (5) (113) and Upilex (6) resin (114) are sold as films, partially polymerized resin coating solutions, and sinterable powders. Allied Signal also offers Apical polyimide film.



(5)

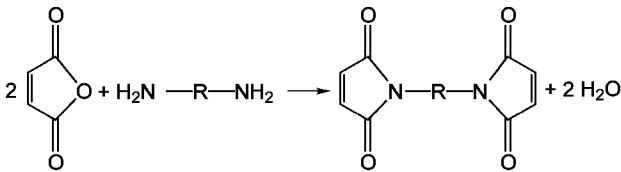


(6)

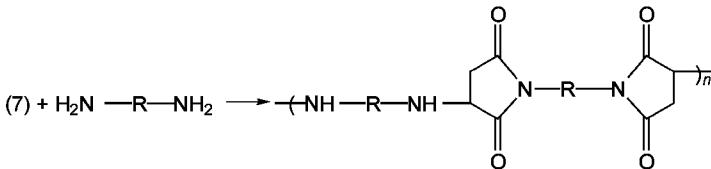
These resins have extensive applications in the high performance insulation markets (traction motor, wire cable) and flexible printed circuits. Du Pont supplies Kapton film and also supplies Verpel stock shapes made from the same resin by powder sintering methods.

Thermoplastic Condensation Polyimides. These include General Electric's Ultem Resin and Amoco's Tordon polyamideimide although the latter is no longer offered as injection moldable pellets, but as compression moldable powder and in solution. Another resin, P-84, originally developed by Upjohn but now made by Lenzing (Austria), is based on benzoquinone dianhydride and aromatic isocyanates.

Additive Polyimides. Rhône-Poulenc's Kinel molding compound and Kerimid impregnating resin (115), Mitsubishi's BT Resins (116), and Toshiba's Imidaloy Resin (117) are based on bismaleimide (4) technology. Maleic anhydride reacts with a diamine to produce a diimide oligomer (7). Further reaction with additional diamine (Michael addition) yields polyaminobismaleimide prepolymer with terminal maleic anhydride double bonds. Cure is achieved by free-radical polymerization through the terminal double bonds.



(7)



Although these resins have been on the market for a number of years, volumes and applications remain low in spite of favorable properties including high temperature resistance, excellent wear and friction properties, good electrical properties, chemical inertness, and inherent nonflammability. Processibility and price may be the limiting factors.

Other Polyimides. In 1979, Rohm & Haas introduced Kamax resin, which was thought to be an *N*-methylamine imidization product of poly(methyl methacrylate) (118). The product was then withdrawn, but was reintroduced in the late 1980s. The partly imidized resins are similar to poly(methyl methacrylate) but have a higher glass-transition temperature.

The total production volume of these various polyimides is low (<1000 t/yr). Prices tend to be high; for example, some Kapton film topics can run in excess of \$176/kg.

Specialty Polyolefins

As recently as 1986 almost all addition polymers were excluded from the ranks of engineering plastics. However, progress since then has been made in the development of addition polymeric resins such as polymethylpentene and polycyclopentadiene and its copolymers (see CYCLOPENTADIENE AND DICYCLOPENTADIENE).

Blends and Alloys

The differentiation between polymer blends (qv) and polymer alloys is blurred. The terms (often used interchangeably), refer to multiresin compositions. Such systems have become increasingly important. For thermodynamic reasons (low entropy of mixing), polymers are usually not miscible with each other; mutual solubility in all proportions is the exception rather than the rule. Mutually compatible mixtures are often referred to as alloys to distinguish them from others where the component resins exist as separate entities in well-defined if random geometrical arrangements, with good adhesion between them. These latter systems are more frequently referred to as blends. Adhesion between the phases, essential if the system is to possess mechanical integrity, is achieved by physicochemical means. Compatibilization, as the process is called, serves to decrease the high interfacial energy between the two immiscible resins. The linkages are effected by grafting, surfactant technology, direct covalent bonding, ionic bonding, and filler penetration.

Alloys exhibit physical properties, the values of which are typically the weighted average of those of its constituents. In particular, the blend exhibits a single glass-transition temperature, often closely obeying semitheoretically derived equations. Blends of two compatibilized immiscible polymers exhibit physical properties which depend on the physical arrangement of the constituents and thus may be much closer to those of one of the parent resins. They will also typically exhibit the two glass-transition temperatures of their constituent resins.

Resin compatibility is not a simple concept. For example, PET–PC blends can be either compatible or noncompatible, depending on stoichiometry and molecular weights. Blends of these two resins in which the PC resin concentration is less than 30 wt % have one glass-transition temperature. On the other hand, a 50–50 mixture of the two resins, although seemingly soluble, retains the glass-transition temperatures of each of the neat resin components. In this compositional range, the resins are incompatible (119). In certain cases of controlled morphology blends, such as the Du Pont Selar resin, laminar inclusions of one resin in another is, in fact, sought (120).

Grafting is employed when one of the resins can react to produce end groups or side-chain functionalities which are soluble or provide a bonding interaction with the other blend resin. The grafted resin can be a primary blend resin, but more often it is used as a blend stabilizer. Grafted PET is an example. Graft points are typically copolymerized acrylate esters. The grafted resins thus have reactive vinyl terminating groups which react with styrenic, rubber, and similar moieties.

Heat-stable, nonvolatile surfactants facilitate processing. Resins with surfactant properties, such as styrene maleic anhydride resins, have sometimes been employed (95).

Direct covalent bonding or ionic bonding can be achieved by inducing chain rupture and chemical combination through shear forces in blending extruders or by the addition of peroxide cross-linking agents. This method resembles grafting. Fillers (qv) stabilize blends by retarding segregation.

Alloys and blends are of great commercial significance. The archetype of "alloys" is the poly(phenylene oxide)–polystyrene resin discussed earlier. Important examples of blends based on immiscible resins are afforded by the polycarbonate–poly(butylene terephthalate) resins and polycarbonate–ABS blends.

Engineering resins can be combined with either other engineering resins or commodity resins. Some commercially successful blends of engineering resins with other engineering resins include poly(butylene terephthalate)–poly(ethylene terephthalate), polycarbonate–poly(butylene terephthalate), polycarbonate–poly(ethylene terephthalate), polysulfone–poly(ethylene terephthalate), and poly(phenylene oxide)–nylon. Commercial blends of engineering resins with other resins include modified poly(butylene terephthalate), polycarbonate–ABS, polycarbonate–styrene maleic anhydride, poly(phenylene oxide)–polystyrene, and nylon–polyethylene.

Companies produce alloys and blends in order to extend their existing resin capacity and product lines without high capital investment and to achieve property profiles required for specific applications.

The properties of PBT and PC resins and of a blend of these two resins are given in Table 18. The chemical resistance of crystalline PBT is reduced, but that of amorphous PC is increased. Hydrolytic stability is good throughout. Impact performance is lower than that of the components. It can be improved by modifiers. A commercial example of this type of resin blend is the General Electric Xenoy resin which is used in automotive bumpers.

Table 18. Properties of Polycarbonate–Poly(butylene terephthalate) Blends^a

Property	ASTM method	PBT	50–50 blend	PC
HDT at 1.82 N/mm ^{2b} , °C	D648	130	99	132
notched Izod at 20°C, 3.2-mm thickness, J/m ^c	D256	50	25	800
flex modulus at 20°C, N/mm ^{2b}	D790	2300	2000	2300
specific gravity	D792	1.31	1.25	1.21
organic solvent resistance		good	fair	poor

^a Ref. 121.
^b To convert N/mm² to psi, multiply by 145.
^c To convert J/m to ftlb/in., divide by 53.38.

Modified nylons are blends of nylon resins and specially grafted nylon resins. In the Du Pont family of Zytel resin, certain blends have been designated Supertough to emphasize the improvement in impact that blends provide over standard resins. General Electric's Noryl GTX resins consist of a nylon matrix resin and a PPO resin in dispersed form. A highly sophisticated blend, it maintains a filled nylon's HDT with no sacrifice of impact resistance.

Blends of PC and ABS resins are being used more and more. Sufficient PC is used to raise the heat-distortion temperature of the ABS above 100°C. Impact resistance is reduced without sacrificing utility, and cost saving can be considerable.

Currently, over 110,000 t/yr of engineering resin blends are consumed worldwide, primarily in the transportation, business-machine, hardware, electrical, and appliance industries. Annual growth is projected to be ca 17% /yr. New blends based on PC, terephthalate, and nylon resins are experiencing the greatest expansion (122). These projections could be surpassed if large-volume metal applications such as automotive panels are replaced by engineering resin blends which are currently being field-tested.

Outlook

The forecasts made in 1985 (77) of 8–8.5% worldwide annual growth have not materialized. The 2 × 10⁶ /yr engineering plastic production reported for 1985–1986 has remained fairly constant. Whereas some resins such as PET, nylon-6, and nylon-6,6 have continued to experience growth, other resins such as poly(phenylene oxide) have experienced downturns. This is due to successful inroads from traditional materials (wood, glass, ceramics, and metals) which are experiencing a rebound in applications driven by new technology and antiplastics environmental concerns. Also, recycling is likely to impact production of all plastics.

The engineering plastics product mix is, however, volume considerations aside, likely to change. Current resins will most likely be modified, as opposed to the marketing of new resins owing to investment considerations. Modification of resin molecular weights, branching and endcapping, compounding with impact modifiers, flame retardants, stabilizers, and reinforcements, and blending can be expected (123). Nylons (124) will offer chemical and heat resistance, modulus, and toughness in order to replace metals in large parts. Blends will afford engineering properties at lower prices. Liquid-crystal (79), advanced polyimide, and amorphous-nylon resin technologies (60) will be applied.

Processing will continue to employ injection molding; foam injection molding and extrusion to produce profiled stock, sheet, and film; blow molding, coextrusion blow molding, and multilayer extrusion will be utilized. Large parts, such as those for the fast-growing automotive exterior markets (125), are likely applications. Graphite-reinforced engineering plastic composites are commercially available in PEEK and polysulfone (26). Additional uses

are expected in applications now being served by reinforced thermoset composite materials (126). Much activity will be seen in reaction-injection-molded nylons, with large parts being obvious targets (127). Conductive plastics are being developed especially for parts requiring inherent r-f shielding capability (128).

Engineering-resin suppliers will work with manufacturers using computer-aided design, engineering, and manufacturing methods (129). The time between product conception and marketing should be dramatically reduced.

BIBLIOGRAPHY

"Engineering Plastics" in *ECT* 3rd ed., Vol. 9, pp. 118–137, by M. Bakker, Business Communications Co., Inc.

1. B. F. Greck, *Chem. Eng. News* **10**, 54 (June 1991).
2. D. N. Lapedes, ed., *Dictionary of Scientific and Technical Terms*, 2nd ed., McGraw-Hill, Inc., New York, 1978, p. 542.
3. D. W. Fox and E. N. Peters in R. W. Tess and G. W. Poehlein, eds., *Applied Polymer Science*, 2nd ed., The American Chemical Society, Washington, D.C., 1985, pp. 495–514.
4. G. F. Foy, *Engineering Plastics and Their Commercial Development*, in R. F. Gould, ed., *Advances in Chemistry Series*, The American Chemical Society, Washington, D.C., 1969, pp. 3 and 4.
5. O. H. Fenner in C. A. Harper, ed., *Handbook of Plastics and Elastomers*, McGraw-Hill, Inc., New York, 1975, pp. 4-1–4-34.
6. H. Mark and G. S. Whitby, eds., *Collected Papers of W. H. Carothers*, Interscience Publishers, New York, 1940.
7. **U.S. Pat. 3,875,116** (Apr. 1, 1975), D. R. Heath and J. G. Wirth (to General Electric Co.).
8. C. W. Bunn in R. Hill, ed., *Fibers from Synthetic Polymers*, Elsevier, Amsterdam, the Netherlands, 1953, pp. 240–256.
9. E. M. Lacey in M. I. Kohan, ed., *Nylon Plastics*, John Wiley & Sons, Inc., New York, 1973, p. 268.
10. J. Brandrup and E. H. Immergut, eds., *Polymer Handbook*, 2nd ed., John Wiley & Sons, Inc., New York, 1975, p. VIII-5.
11. D. C. Clagett in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 6, Wiley-Interscience, New York, 1986, p. 96.
12. R. D. Deanin in Ref. 4, pp. 9–12, 18.
13. H. Faig in J. Agranoff, ed., *Modern Plastics Encyclopedia 1982–1983*, Vol. 60, No. 10A, McGraw-Hill, Inc., New York, 1983, pp. 248–262, 268.
14. Ref. 11, p. 103.
15. F. R. Nissel in Ref. 13, pp. 234–237.
16. K. Alex in Ref. 13, pp. 237 and 238.
17. M. Yahr in Ref. 13, pp. 178–188.
18. R. W. Haviland in Ref. 13, pp. 188–190.
19. S. L. Belcher in Ref. 13, pp. 190 and 191.
20. E. Steward in Ref. 13, pp. 214–231.
21. W. F. Kent in Ref. 13, pp. 322–327.
22. C. M. Hall in Ref. 13, p. 311.
23. P. E. Gage in Ref. 13, pp. 193 and 194.
24. R. W. Lenz, *Organic Chemistry of Synthetic High Polymers*, Interscience Publishers, New York, 1967, p. 562.
25. R. P. Geer, *Plast. Eng.* **41**(11), 41–49 (1985); D. S. Breslow, *Chem Tech* **20**, 540–544 (Sept. 1990).
26. R. B. Rigby, *VICTREX Polyethersulfone (PES) and VICTREX Polyetherethersulfone (PEEK)*, SPE-RETEC, Mississauga, Ontario, Canada, Oct. 17, 1984, pp. 6 and 7.
27. S. G. Hill, E. E. House, and J. T. Haggatt, *Final Report Advanced Thermoplastic Composite Development Contract N 00019-77-C-0561*, Naval Air Systems Command, U.S. Department of the Navy, Washington, D.C., May 1979, pp. 7–9.
28. *Annual Book of ASTM Standards*, The American Society for Testing and Materials, Philadelphia, Pa., 1984, Index.
29. *Tests for Flammability of Plastic Materials for Parts in Devices and Appliances*, UL 94, 3rd ed., Underwriters Laboratories, Inc., Northbrook, Ill., 1980.
30. Ref. 11, p. 100.
31. C. A. Harper in Ref. 5, Chapt. 2.
32. G. E. Rudd and R. W. Sampson in Ref. 5, Chapt. 3.
33. R. S. Lenk, *Polymer Rheology*, Applied Science Publishers, Ltd., London, 1978.
34. R. M. Bonner and co-authors in Ref. 9, p. 337.
35. C. A. Harper in Ref. 5, p. 347.
36. Ref. 28, D1925, 08.02.
37. *Polymeric Materials—Long Term Property Evaluations*, UL 746B, 2nd ed., Underwriters Laboratories, Inc., Northbrook, Ill., 1979, pp. 6–26.
38. R. L. Bergen, Jr., *SPE J.* **18**, 667 (1962).
39. P. D. Gabriche, J. R. Geib, R. M. Iannusis, and W. J. Reid in D. P. Garner and G. A. Stahl, eds., *The Effects of Hostile Environments on Coatings and Plastics*, The American Chemical Society, Washington, D.C., 1983, pp. 317–330.
40. S. Collins, *Plast. Des. Forum* **8**(2), 17 (1983).
41. M. Huecker and S. Blackson, *CAE*, 56 (1984).
42. R. L. Maffly, J. Bakker, and K. Tsuchiya in E. Helmes, E. M. Connolly, L. Waddams, and J. Shimosato, eds., *Chemical Economics Handbook*, SRI International, Menlo Park, Calif., 1983, pp. 580.0951 D–I.
43. Ref. 11, p. 121.
44. *DELTRIN General Guide to Products and Properties*, Brochure E-43409, E. I. du Pont de Nemours & Co., Inc., Wilmington, Del., p. 3.
45. *Ibid.*, p. 4.
46. *CELCON Acetal Copolymer Properties*, Brochure 10M, Celanese Corp., New York, June 1983.
47. A. Bettelheim, *Plast. Technol.* **31**, 20, 22 (1985).
48. H. E. Grey, R. L. Maffly, E. Helenes, J. Bakker, and J. Shimosato in Ref. 42, 1983, pp. 580.0821 C.
49. R. M. Bonner and co-workers in Ref. 9, pp. 399 and 400.
50. L. T. Sherwood in Ref. 13, pp. 28–34.
51. H. Doffin, W. Pungs, and R. Gabler, *Kunststoffe* **56**, 542 (1966).
52. H. E. Grey, R. L. Maffly, E. Helenes, J. Bakker, and J. Shimosato in Ref. 44, 1983, pp. 580.0822 D–E, Q–S, V.
53. Ref. 24, pp. 103–106.
54. J. March, *Advanced Organic Chemistry*, 2nd ed., McGraw-Hill, Inc., New York, 1977, p. 1171.
55. P. Kovacic and F. W. Koch, *J. Org. Chem.* **28**, 1964 (1963).
56. Ref. 24, pp. 106 and 107.
57. G. M. Van der Want and C. A. Knuissink, *J. Polym. Sci.* **35**, 119 (1959).
58. Ref. 11, p. 123.
59. *Designing with Plastics*, Brochure E-32986, E. I. du Pont de Nemours & Co., Inc., Wilmington, Del., pp. 18 and 19.
60. D. Bremer, *Plast. Des. Forum* **8**(1), 58 (Jan.–Feb. 1983).
61. *Ibid.*, pp. 57–64.

62. *Mod. Plast.* **60**, 69, 70 (Apr. 1983).

63. H. E. Grey, R. L. Maffly, E. Helenes, J. Bakker, and J. Shimosato in Ref. 42, 1983, pp. 580.0821 E, G. H.

64. Ref. 110, p. 57.

65. C. Nitschke in Ref. 13, pp. 46 and 47.

66. C. S. Hughes, J. Bakker, and K. Tsuchiya in Ref. 42, pp. 580.1172 I–J, T, W, Z, 580.1173 Q, T, V, W, 580.1174 C, D, F.

67. W. Griehl and G. Schnock, *J. Polym. Sci.* **30**, 413 (1958).

68. C. S. Hughes, J. Bakker, and K. Tsuchiya in Ref. 42, pp. 580.1172 D, F.

69. J. Fredos, *Plast. Focus* **16**(32), 1 (1984).

70. C. S. Hughes, J. Bakker, and K. Tsuchiya in Ref. 42, pp. 580.1172 K–L.

71. J. Antonopoulos in Ref. 13, Vol. 61, No. 10A, 1984, pp. 192–194.

72. C. S. Hughes, J. Bakker, and K. Tsuchiya in Ref. 42, pp. 580.1172 B, F–G, Q.

73. *Ibid.*, pp. 580.1172 C–D, G–I.

74. Ref. 11, p. 115.

75. *VALOX Injection Molding, Brochure VAL15C*, General Electric Co., Pittsfield, Mass., pp. 23–25.

76. *RYNITE General Guide to Products and Properties, Brochure E-36602*, E. I. du Pont de Nemours & Co., Inc., Wilmington, Del., pp. 3 and 4.

77. J. Fredos, *Plast. Focus* **16**(49), 1, 2 (1985).

78. *Chem. Week* **13**(4), 12, 13 (1985).

79. *Br. Polym. Rubber*, 25 (July Aug. 1985).

80. D. Bellius, *Adv. Photochem.* **8**, 109 (1971).

81. Ref. 11, p. 117.

82. W. F. Christopher and D. W. Fox, *Polycarbonates*, Reinhold Publishing Corp., New York, 1962, pp. 161–173.

83. *Plast. Brief* **10**(51), 1 (1981).

84. D. C. Clagett and S. J. Shafer, *Polycarbonate Resins*, SPE-RETEC, Mississauga, Ontario, Canada, Oct. 16, 1984, p. 2.

85. *Ibid.*, p. 3.

86. Ref. 11, p. 119.

87. *Plast. World, 1984 Plast. Directory Plast. Prop. Guide* **42**(4), 54 (1984).

88. *LEXAN Resin Competitive Data, Brochure CDC 672*, General Electric Co., Pittsfield, Mass., p. 8.

89. R. L. Maffly, J. Bakker, and K. Tsuchiya in Ref. 42, 1983, p. 580.1121 C.

90. *Ibid.*, p. 580.1121 D.

91. R. I. Warren, *Polyphenylene Ethers and the Alloys*, SPE-RETEC, Mississauga, Ontario, Canada, Oct. 16, 1984, pp. 1–7.

92. Ref. 11, p. 125.

93. *NORYL Design, Brochure CDX-83*, General Electric Co., Pittsfield, Mass., 1983, pp. 6 and 7.

94. *PREVEX Polymers, Technical Publications PR-102A*, Borg-Warner Chemicals, Inc., Parkersburg, W. Va., 1988.

95. **U.S. Pat. 4,315,086** (Feb. 9, 1982), H. K. Uemo and T. T. Marugama (to Sumitomo Chemical Co.).

96. R. L. Maffly, O. Kamatari, and J. Bakker in Ref. 42, 1982, pp. 580.0600 G, I, J.

97. R. N. Johnson, A. G. Farnham, R. A. Clendinning, W. F. Hale, and C. N. Merriam, *J. Appl. Polym. Sci. Part A-1* **5**, 2375 (1967).

98. Ref. 11, p. 110.

99. *ULTEM, Polyetherimide Properties Guide*, General Electric Co., Pittsfield, Mass., Mar. 1985.

100. **U. S. Pat. 3,354,129** (Nov. 21, 1967), J. T. Edmonds, Jr. and H. W. Hill, Jr. (to Phillips Petroleum Co.).

101. Ref. 11, p. 111.

102. *RYTON Polyphenylenesulfide Physical, Chemical, Thermal, and Electrical Properties, TSM-266*, Phillips Chemical Co., Bartlesville, Okla., Apr. 1981, p. 2.

103. Ref. 11, p. 104.

104. Ref. 26, pp. 107.

105. *Mod. Plast. Int.* **11**(11), 76 (1981).

106. *Chem. Week* **12**, 90 (June 27, 1984).

107. *TORLON, Brochure D0881*, Amoco Chemicals Corp., Chicago, Ill.

108. *TORLON, Technical Data and Applications Information, Brochure 1117-7.5M-480*, Amoco Chemicals Corp., Chicago, Ill.

109. Ref. 42, 1984, p. 580.1401 G.

110. Ref. 11, p. 105.

111. *Amoco TORLON Engineering Resins, Data Sheet Grade 4203L*, Code Number AT-5, Chicago, Ill., Feb. 1982.

112. *Amoco TORLON Engineering Resins, Data Sheet Grade 7130*, Code Number AT-14, Chicago, Ill., Feb. 1982.

113. Ref. 42, p. 580.1401E.

114. *Mod. Plast. Int.* **14**(4), 81 (Apr. 1984).

115. *Resine Polyimide pour Composites Haute Temperatures, Brochure CF H711*, Rhône-Poulenc, Paris (in French).

116. *Plast. Ind. News* **27**(11), 164 (Nov. 1981).

117. Ref. 42, p. 480.1401 V.

118. *KAMAX 201, Engineering Thermoplastic Resin, Product Brochure*, Rohm & Haas Co., Philadelphia, Pa., 1979.

119. T. R. Nasson, D. R. Paul, and J. W. Barlow, *J. Appl. Polym. Sci.* **23**, 85 (1979).

120. P. M. Subramunian, *Permeability Barriers by Controlled Morphology of Polymer Blends*, SPE-RETEC, Mississauga, Ontario, Canada, Oct. 16, 1984, pp. 1–9.

121. Ref. 11, p. 127.

122. M. Rappaport, *Engineering Blends*, SPE-RETEC, Mississauga, Ontario, Canada, Oct. 16, 1984, p. 12.

123. A. S. Wood, *Mod. Plast. Int.* **14**(10), 43 (Oct. 1984).

124. R. Wehrenberg, *Plast. World* **42**, 39 (Sept. 1984).

125. *Plast. World* **42**, 32 (Oct. 1984).

126. D. McCall and Briefing Panel, *Report of the Research Briefing Panel on High Performance Polymers and Polymeric Composites*, National Research Council, Washington, D.C., Aug. 10, 1984, pp. 11–12, 16–19.

127. R. Wehrenberg, *Plast. World* **42**, 62 (Oct. 1984).

128. A. Sternfield, *Mod. Plast. Int.* **12**(4), 48 (July 1982).

129. *Plast. Rubber Week.*, 6 (July 21, 1984).

Donald Clagett
Provisional Technical Secretariat Organisation for the Prohibition of Chemical Weapons

Raymond Narr
General Electric Company

ENHANCED RECOVERY.

See PETROLEUM.

ENKEPHALINS.

See HORMONES, BRAIN OLIGOPEPTIDES; NEUROREGULATORS; OPIOIDS, ENDOGENOUS.

ENVIRONMENTAL IMPACT

The significant impact of environmental regulations in the 1990s on industry in general, and the chemical industry in particular, is indisputable. Public awareness of the need to avoid endangerment of the natural systems that support life on earth, ie, the atmosphere (see AIR POLLUTION; AIR POLLUTION CONTROL METHODS; ATMOSPHERIC MODELING), water (qv), soil, and living things, has also intensified.

A concomitant increase in the coverage of environmental issues is evidenced in the *Encyclopedia*. There are new articles, ie, GROUNDWATER MONITORING, HAZARD ANALYSIS AND RISK ASSESSMENT, HAZARDOUS WASTE TREATMENT, and WASTE REDUCTION, which have environmental topics as the primary focus. There is also a growing tendency to merge environmental issues and economics in decision making within the chemical industry (see CATALYSTS, REGENERATION; ECONOMIC EVALUATION; FUELS FROM WASTE; HERBICIDES; RECYCLING; SOLVENTS, INDUSTRIAL; WASTES, INDUSTRIAL). Moreover, because environmental issues are pervasive in the chemical industry, these concerns have brought about change in chemical technology. In many fields, environmental concerns and regulations are driving technological development (see COATING PROCESSES; DYES, ENVIRONMENTAL CHEMISTRY; EXHAUST CONTROL, AUTOMOTIVE; EXHAUST CONTROL, INDUSTRIAL).

The impact of environmental issues is apparent within the majority of *Encyclopedia* articles having counterparts in previous editions. For example, the effect of environmental concerns on manufacturing and other processes is evident in articles such as ALKALI AND CHLORINE PRODUCTS, ELECTROPLATING, MINERAL RECOVERY AND PROCESSING, and PETROLEUM. Concern about the environment has also played a role in the types and quantities of materials produced (see CHLOROCARBONS AND CHLOROHYDROCARBONS; COATINGS, MARINE; CORROSION AND CORROSION CONTROL; PIGMENTS, INORGANIC) as well as with regard to the technology employed (see COAL CONVERSION PROCESSES; FOOD PACKAGING).

ENZYME APPLICATIONS

Industrial,
Therapeutic,

INDUSTRIAL

Enzymes, like other proteins (qv), are composed of up to 20 different amino acids (qv). They accelerate hundreds of reactions taking place simultaneously in the cell and its immediate surroundings, and are essential for the development and maintenance of life.

Industrial applications of enzymology form an important branch of biotechnology. Enzymatic processes enable natural raw materials to be upgraded and turned into finished products. They offer alternative ways of making products previously made only by conventional chemical processes.

The detergent industry is the largest user of industrial enzymes. The starch industry, the first significant user of enzymes, developed special syrups that could not be made by means of conventional chemical hydrolysis. These were the first products made entirely by enzymatic processes. Materials such as textiles and leather can be produced in a more rational way when using enzyme technology. Foodstuffs and components of animal feed can be produced by enzymatic processes that require less energy, less equipment, or fewer chemicals compared with traditional techniques.

The development of the submerged fermentation technique resulted in tremendous progress in the field of industrial enzymology. This technique was used in the early 1950s for the production of bacterial amylases for the textile industry.

History

In 1833 an amylase from germinating barley was recovered and called diastase (1). Like malt itself, this product converted gelatinized starch into sugars, primarily maltose. Shortly thereafter, Berzelius proclaimed the existence of non-living catalysts, and Schwann (2) reported on his observation and purification of pepsin.

Throughout the second half of the nineteenth century, several schools of thought debated the connection between biological catalysis and life or vital forces, ie, *vis vitalis*. The Pasteur school firmly believed that alcoholic fermentation was a vital act which could not take place without the presence of viable organisms (3). Another school (4) was convinced fermentation was the result of a common chemical process, and that yeast was a nonviable substance continuously engaged in the process of breaking down other substances. A final group of scientists strongly supported the original concept of Berzelius that enzymes and living microorganisms, at that time known collectively as unorganized ferments or simply ferments, were two very different phenomena (5–7).

In 1878 the term enzyme, Greek for "in yeast," was proposed (8). It was reasoned that chemical compounds capable of catalysis, ie, ptyalin (amylase from saliva), pepsin, and others, should not be called ferments, as this term was already in use for yeast cells and other organisms. However, proof was not given for the actual existence of enzymes. Finally, in 1897, it was demonstrated that cell-free yeast extract ("zymase") could convert glucose into ethanol and carbon dioxide in exactly the same way as viable yeast cells. It took some time before these experiments and deductions were completely understood and accepted by the scientific community.

Early Industrial Enzymes. Enzymes were used in ancient Greece for the production of cheese (9). Early references to this are found in Greek epic poems dating from about 800 BC. Fermentation processes for brewing, baking, and the production of alcohol have been known since prehistoric times.

The era of modern enzyme technology began in 1874 when the Danish chemist Christian Hansen produced the first industrial batches of chymosin by extracting dried calves' stomachs with saline solutions.

One of the first large-scale industrial productions and applications of enzyme technology to emerge in the twentieth century was the production of fungal amylases by the surface or semisolid culture fermentation of *Aspergillus oryzae* on moist rice or wheat bran. This process, developed by the Japanese scientist Takamene, was inspired by traditional koji fermentation in trays used for the production of foodstuffs and flavoring based on soy protein; this method has been largely replaced by the more efficient submerged fermentation. Takamene's product was called Takadiastase, and is still in use as a digestive aid.

Textile and Leather Industries. At about the same time Takamene was developing his fermentation technique, enzymes were introduced in the desizing of textiles, ie, the process by which all the starch paste, which has served as strengthening agent and lubricant to prevent breaking of the warp during the weaving process, is removed from the fabric. Historically, textiles were treated with acid, alkali, or oxidizing agent, or soaked in water for several days so that naturally occurring microorganisms could break down the starch. These methods were difficult to control, and sometimes damaged or discolored the material. The application of crude enzyme extracts, from malt or pancreatic glands, in desizing was a significant step forward. The next development was the introduction, on a small scale, of a bacterial amylase from *Bacillus amyloliquefaciens* by Boidin and Effront in 1917. Mass production did not begin until after World War II, when submerged fermentation was developed as a substitute for the original surface fermentation.

The German chemist and industrial magnate Dr. Otto Rühm, founder and partner of Rühm and Haas in Darmstadt, Germany, was responsible for the further development of industrial enzymes. He studied the leather bating process, ie, the operation in the processing of hides and skins that precedes the tanning step. Bating removes some of the protein that is not essential for the strength of the leather (qv) and that might otherwise prevent the leather from achieving the suppleness and softness of touch required in numerous products. It serves to control the quality of leather; eg, stiff leather used for soles is only lightly bated, but the soft qualities required for gloves result from intense bating. The traditional bating process used dog and pigeon excrements. Rühm's theory was that digestive enzymes were responsible for the bating process, and he correctly concluded that extracts of the pancreas might be used directly for bating.

Detergent Industry. In 1913, Dr. Rühm assumed that dirt in fabrics used by humans was to a large degree composed of fats and protein residues and that the well-known ability of tryptic enzymes to break down fat and protein might be exploited in laundry cleaning (10); trypsin was subsequently added to the wash (11). The concept was immediately commercialized by Rühm and Haas in the presoak detergent Burnus, consisting of soda and small amounts of pancreatin. It sold for the following 50 years in Europe, although it was never a commercial success.

During World War II, a severe shortage of fats and soap inspired the development of another enzymatic presoaking agent, Bio 38 (Gebrüder Schnyder). Some years later, Bio 40, a product containing a bacterial protease, was launched onto the market. This protease was a considerable improvement over the previous pancreatic products, but still suffered from the disadvantage of a neutral pH optimum, giving it low activity and stability in a washing solution.

In 1958, the microbial alkaline protease Alcalase (Novo Industries) was produced by fermentation of a strain of *Bacillus licheniformis*. It had high stability and activity at pH 8–10, was marketed in 1961, and was incorporated into Bio 40. However, it was not until the successful marketing of the presoaking agent Biotex in 1963 that detergent manufacturers saw the true possibilities of enzymes.

Several new detergent enzymes have emerged on the market (Table 1). Truly alkaline proteases, introduced in 1974 and 1982, were fermented on strains of *Bacillus lentus/firmus*. These enzymes have a pH optimum between 9 and 11, and have taken important market shares from Alcalase.

Table 1. Available Industrial Enzymes, 1993

Year launched	Enzyme type	Brand names ^a	Company
---------------	-------------	--------------------------	---------

1950	bacterial α -amylase	BAN	Novo Nordisk
		Canalpha	Solvay Enzymes
		Rapidase	IBIS
1963	protease with a low alkaline pH optimum	Alcalase	Novo Nordisk
		Maxatase	IBIS
		Optimase	Solvay Enzymes
		Superase	Pfizer
		Milezyme	Solvay Enzymes
1965	amyloglucosidase	AMG	Novo Nordisk
		Amigase	IBIS
		Optidex	Solvay Enzymes
		Optilase	Solvay Enzymes
		Diazyme	Solvay Enzymes
		Spezyme	Genencor International
		Termamyl	Novo Nordisk
1973	thermostable α -amylase	Maxamyl	IBIS
		Takatherm	Solvay Enzymes
		Optitherm	Solvay Enzymes
		Kleistase	Daiwa Kasei
		Rennilase	Novo Nordisk
1973	microbial rennet (chymosin)	Esperase	Novo Nordisk
1974	protease with high alkaline pH optimum	Sweetzyme	Novo Nordisk
1976	glucose isomerase	Maxazym	IBIS
		Optisweet	Solvay Enzymes
		Spezyme	Pfizer
		Spezyme	Genencor International
		Takasweet	Solvay Enzymes
		Savinase	Novo Nordisk
		Maxacal	IBIS
		Opticlean	Solvay Enzymes
		Thermozyme	Novo Nordisk
		Celluzyme	Novo Nordisk
1982	low temperature protease with high alkaline pH optimum	Lipolase	Novo Nordisk
1982		Maxapem	IBIS
1986		Durazym	Novo Nordisk
1988			
1991			
	bleach-stable protease with high pH optimum made by protein engineering		

^a First product on market is listed first for each enzyme type.

In 1989, two enzymes based on genetic engineering techniques were introduced, ie, a cloned alkaline protease (IBIS) and a protein engineered Subtilisin Novo (Genencor, California). Lipase and cellulase types of detergent enzymes have also begun to appear.

Starch Industry. Enzymes have been of great value to the starch industry since the 1960s. In the 1950s, fungal amylases were used in the manufacture of specific types of syrups. Early in the 1960s, the enzyme amyloglucosidase made it possible to completely break down starch into glucose. A few years later, most glucose production switched from the old acid hydrolysis method to the enzymatic process, which gave better yields, higher purity, and easier crystallization. This process was further improved by the introduction of a new technique used for the enzymatic pretreatment (liquefaction) of starch. A new and extremely thermostable bacterial α -amylase was introduced in 1973.

In the 1980s the focus was on glucose isomerase, which converts glucose into fructose and thereby doubles the sweetness of the sugar. The first enzyme of this type for the industrial market was launched in 1976 under the name Sweetzyme (Novo). The combination of this product with amyloglucosidase and α -amylase made it possible to use starch as a raw material for production of a sweetening agent with almost the same composition, amount of calories, and sweetening effect as ordinary cane or beet sugar. The new product, high fructose corn syrup (HFCS) or isosyrup, dominates the market for sweeteners (qv) (ca 1993).

Other Industrial Enzymes. Many new industrial applications of enzymes have been developed since the 1960s. Significant areas of applications include fungal proteases as substitutes for calf chymosin in cheesemaking, enzymes for brewing and for the wine (qv) and juice industries, and enzymes for the baking industry (see BAKERY PROCESSES AND LEAVENING AGENTS; BEER). Other areas, like paper (qv), pulp (qv), and bioremediation, are working intensively at implementing enzyme technology (see HAZARDOUS WASTE TREATMENT).

Genetic Engineering of Enzymes

With the usefulness of the first industrial enzymes, scientists and companies began to look for sources of new and better enzymes. Originally, attention was focused on animal and plant material, ie, the pancreas and malt. Although these could well serve as cheap raw materials, it became clear in the 1950s and 1960s that enzymes of microbial origin offered greater advantages; eg, microbes can be grown quickly and efficiently in fermentation tanks, and natural productivity of an enzyme can be increased by mutating microbes and screening for higher yielding mutant strains. Intense efforts were made to screen nature, in particular soil samples, for microbes producing enzymes that were suitable for industrial development. As a consequence, two dozen new enzymes were marketed within a few decades. Most of these production organisms belong either to the genus *Bacillus* (gram-positive bacteria) or *Aspergillus* (filamentous fungi) and most enzymes are either amylases or proteases. Amylases and proteases are enzymes that are usually secreted, ie, released from the cells into the growth medium. Therefore, the likelihood of high yields and simple purification procedures are increased. Since many *Aspergilli* and *Bacilli* have a characteristic ability to secrete amylases, proteases, and other enzymes, they were a natural target group for screening. Many of these species are easily isolated, are robust to handle under laboratory conditions, and grow well in fermentors on cheap industrial media to high cell densities.

Interesting enzymes were also detected in other types of microorganisms. Further development of most of these enzymes had to be given up for one or more reasons, ie, it was impossible to grow the microorganism successfully under industrial conditions; the microorganism was pathogenic, toxinogenic, or not completely safe to handle; enzyme purification was prohibitively expensive because, for instance, the enzyme was cell associated or contaminated with undesirable enzymes or other compounds; and yields of the desired enzyme could not be increased by classical mutagenesis and screening for higher yielding strains. Genetic engineering now offers new solutions to these problems.

Genetic engineering makes it possible to take DNA out of a donor cell, modify the DNA in a test tube, and introduce the modified recombinant DNA into a new host cell in such a way that the recombinant DNA becomes a stable part of the genetic material of the host. Thus it is possible to transfer the gene encoding a particular enzyme from any exotic organism into a well-known production organism like a *Bacillus* or an *Aspergillus*; ie, a production organism may be obtained that both produces the desired enzyme of the exotic donor and has all the desirable properties of a safe industrial microorganism.

The process of isolating a particular gene by genetic engineering techniques is commonly referred to as molecular cloning or simply cloning (see GENETIC ENGINEERING). The efficiency of expression of a gene, ie, the amount of protein produced, depends on a number of properties of both the gene itself and the host organism. For industrial purposes, enzyme expression has to be very high, usually several grams per liter of fermentation liquid. To

achieve this, it may be sufficient to choose a suitable host that resembles the donor organism. However, in most cases modifications by genetic engineering are necessary to increase the activity of the gene. Even then, combination of a suitable host with a gene designed to be very active may not be sufficient. One reason for this may be that the enzyme cannot be secreted, ie, transported across the cell membrane and cell wall of the host into the medium.

Almost all important industrial enzymes are naturally secreted. The substrates of these enzymes are large polymeric molecules such as proteins (qv), starch (qv), and cellulose (qv) that cannot be taken up by the cell and digested by internal enzymes. Consequently, the cell must secrete appropriate enzymes to utilize the chemical energy of the polymers. The cell does this by attaching to the enzyme a string of amino acids called a signal peptide. The signal peptide is recognized by and bound to the cell membrane, whereby the entire enzyme is attached to the inner surface of the membrane. While the signal peptide serves as an anchor, the enzyme is pulled through the membrane and subsequently released on its outer surface following removal of the signal peptide. A particular signal peptide usually works well with its natural enzyme in its natural host. When cloned in another host, the yield of the enzyme can be increased by replacing the native signal peptide with one known to function well in the new host. If the enzyme is naturally intercellular, it may not be possible to obtain secretion with any signal peptide because the enzyme rapidly folds in a way that prevents secretion. Thus the ability to be secreted also depends on the properties of the enzyme itself.

Host Microorganisms. As of this writing, only microorganisms are used as recombinant production organisms for industrial enzymes; therefore, tissue cultures and transgenic plants and animals are not dealt with herein. The choice of host microorganism for production of industrial enzymes is often critical for the commercial success of the product. Potential hosts should give sufficient yields, be able to secrete large amounts of protein, be suitable for industrial fermentations, produce a large cell mass per volume quickly and on cheap media, be considered safe based on historical experience or evaluation by regulatory authorities, and should not produce harmful substances or any other undesirable products. Few microorganisms fulfill all the above criteria and are used for production of industrial enzymes. Important hosts are presented as follows.

Escherichia coli. The genetics of this gram-negative bacterium are very well known. For this reason, many of the first efforts to produce recombinant products from this microorganism were successful. However, because of the importance of the other criteria listed above, many efforts failed. *E. coli* is only used to produce the milk-clotting mammalian protease chymosin [9001-98-3] (rennin).

Bacilli. This very diverse group of gram-positive bacteria includes several historically well-known enzyme producers that fulfill all the above criteria including the ability to secrete large amounts of enzyme. It is generally known that recombinant *B. subtilis*, *B. licheniformis*, and *B. alkalophilus* strains are used for enzyme production.

Aspergilli. The filamentous fungi also include a number of well-known enzyme producers. Among these, *A. oryzae* was used as a host for lipase, the first industrial enzyme produced from a genetically engineered organism and sold on a large scale. *A. niger* is used for the production of chymosin. Recombinant production organisms derived from these two species may be used for production of several enzymes.

Yeast. Several yeast species, including *Saccharomyces cerevisiae* (baker's yeast) and *Khyveromyces lactis*, are good candidates for the production of certain industrial enzymes, although their ability to secrete is much inferior to *Bacilli* and *Aspergilli*. The best-known example of *K. lactis* is used for commercial production of chymosin [9001-98-3].

It is increasingly evident that even when an efficiently expressed gene is inserted into a suitable host, the optimal production strain is rarely obtained. Among the properties that may need to be improved are genetic and product stability.

One method used to achieve genetic stability is to insert the plasmid or the recombinant DNA directly into the chromosome. Since no cell can afford to lose a chromosome, this assures that the recombinant DNA is not lost as long as it remains an integral part of the chromosome.

Microorganisms that secrete proteins often secrete proteases. Many secreted enzymes are actually resistant to proteolytic degradation, at least in their natural host, but some are not. To maintain product stability, it may be necessary to remove or reduce the proteolytic activity of the host. This can be achieved by classical mutagenesis, particularly in those cases where the microorganism only produces one protease, or by replacing normal genes on the chromosome with genes that have been inactivated in the test tube. This precise modification normally does not affect other properties of the host, and proteases may be inactivated one after another. This approach, however, requires that suitable genetic engineering techniques be applied to the particular host strain, and that each of the proteases is cloned.

Engineering of New Enzymes. There are approximately 10 million plant and animal species on Earth, and most of these contain hundreds of enzymes different from the enzymes of any other species. Furthermore, new enzymes are continuously created in nature by evolutionary processes. Sophisticated screening systems are now being developed to find new enzymes in microorganisms growing in hot springs, the deep sea, or other extreme environments. The prospect that a microorganism may not be cultivable no longer represents an insurmountable obstacle to the use of its genes to genetic engineering, and screening for microorganisms may be supplemented by screens for genes.

It is possible to modify and improve natural enzymes with protein engineering. Because the properties of any enzyme are determined by its three-dimensional structure, which in turn is determined by the linear combination of amino acids, it is possible, by a precise and site-specific modification of a gene, to change any amino acid into any other, or design a synthetic gene of any composition, and hence design a protein of any three-dimensional structure as long as basic physical and chemical rules are obeyed. Protein engineers can only predict the structural consequences of minor changes in enzymes (ca 1993). Protein-engineered enzymes on the market include a detergent protease that has been stabilized against chemical oxidation by replacement of two amino acids of the native enzyme.

It is likely that any new enzymes isolated by screeners will be quickly and routinely cloned by genetic engineers, and be sequenced and expressed as almost pure proteins. Protein chemists can then evaluate the properties of the new enzyme and determine its three-dimensional structure. This vast amount of information allows the protein engineers and their computers to design the enzymes of the future.

Abzymes. To design an improved enzyme, protein engineers must know the structure of a protein resembling the desired enzyme. Since the number of different known types of enzymes is less than 1000, only a limited number of chemical reactions can be catalyzed by known naturally occurring enzymes. Catalytic antibodies, or abzymes, may overcome this limitation. Abzymes are antibodies having the ability to complex with the transition state of a chemical reaction, thereby lowering the free energy of the intermediate. Since the reaction rate is determined by the difference in free energy between the reactants and the transition state, the lowering of the energy of the transition state speeds up the process until chemical equilibrium, which is independent of catalysis, is reached. Since the immune system of mammals is able to produce approximately 100 million antibodies, it is likely that an antibody that recognizes any transition-state molecule, or a stable analogue of this, can be raised. The concept has been shown to work with at least seven different types of reactions, including a specific ester hydrolysis which was stimulated one millionfold by an abzyme. A cost-effective industrial abzyme must still be developed.

Catalytic Activity

Enzymatic Catalysis. Enzymes are biological catalysts. They increase the rate of a chemical reaction without undergoing permanent change and without affecting the reaction equilibrium. The thermodynamic approach to the study of a chemical reaction calculates the equilibrium concentrations using the thermodynamic properties of the substrates and products. This approach gives no information about the rate at which the equilibrium is reached. The kinetic approach is concerned with the reaction rates and the factors that determine these, eg, pH, temperature, and presence of a catalyst. Therefore, the kinetic approach is essentially an experimental investigation.

The characteristics of enzymes are their catalytic efficiency and their specificity. Enzymes increase the reaction velocities by factors of at least one million compared to the uncatalyzed reaction. Enzymes are highly specific, and consequently a vast number exist. An enzyme usually catalyzes only one reaction involving only certain substrates. For instance, most enzymes acting on carbohydrates are so specific that even the slightest change in the stereochemical configuration is sufficient to make the enzyme incompatible and unable to effect hydrolysis.

Usually the degree of specificity of an enzyme is related to its biological role. Subtilisins, proteolytic enzymes secreted by certain bacteria, are relatively indiscriminating about the nature of the side chains adjacent to the peptide bond to be cleaved, although these enzymes preferably hydrolyze peptide bonds on the carboxylic side of aromatic or aliphatic nonpolar side chains. This reflects the fact that the function of these enzymes is to hydrolyze larger proteins into smaller peptides that can be absorbed more easily by the microorganism. The digestive enzyme trypsin splits bonds on the carboxylic side of lysine and arginine residues only. The enzymes involved in blood clotting are even more specific than trypsin.

Another characteristic of enzymes is their frequent need for cofactors. A cofactor is a nonprotein compound that combines with the otherwise inactive enzyme to give the active enzyme. Examples of cofactors are metal ions such as Ca^{2+} , Cu^{2+} , Co^{2+} , Fe^{2+} , and Mg^{2+} , and organic molecules such as nicotinamide adenine dinucleotide (NAD) and flavin adenine dinucleotide (FAD).

Enzymes accelerate reactions by stabilizing the transition states, the highest energy species on the reaction pathway, and thereby decreasing the activation barrier. In other words, the combination of enzyme and substrate creates a new reaction pathway whose transition-state energy is lower than it would be if the reaction were taking place without the participation of the enzyme. Enzymes have evolved to bind the transition states of substrates more strongly than the substrates themselves. Therefore, compounds that mimic the structure of the transition state are often potent inhibitors of the enzyme-catalyzed reaction.

Enzyme Kinetics. A simple enzyme catalyzed reaction can be described:



The enzyme, E , and the substrate, S , initially combine to form an enzyme–substrate complex, ES , which is held together by physical forces. In the second step the chemical process occurs, whereby the enzyme and the product, P , are liberated. This step is controlled by a first-order rate constant k_{cat} , called the catalytic constant or the turnover number. When deriving kinetic expressions, it is generally assumed that the concentration of enzyme is negligible compared to the concentration of substrate. Furthermore, it is assumed that what is being measured is the initial velocity v for the formation of products, ie, the rate of formation for the first few percent of the product. Under these conditions the products have not accumulated, the substrates have not been depleted, and the reaction velocity is generally linear with time. It is found experimentally that v is directly proportional to the concentration of enzyme $[E_0]$, but varies with the substrate concentration $[S]$ (Fig. 1). At low $[S]$, v is directly proportional to $[S]$. At higher $[S]$, however, this relation begins to break down, and at sufficiently high $[S]$ v tends toward a limiting value v_{max} . The Michaelis-Menten equation (eq. 2) expresses this relation quantitatively.

$$v = \frac{K_{\text{cat}} [E_0]}{1 + K_m [S]} \quad \text{where } v_{\text{max}} = k_{\text{cat}} [E_0] \tag{2}$$

k_{cat} represents the maximum amount of substrate that the enzyme can convert to products per unit time; for most enzymes this lies in the range 1 to 10^4 s^{-1} . K_m is the Michaelis constant, and is the substrate concentration at which $v = \frac{1}{2} v_{\text{max}}$; K_m for most enzymes lies in the range 10^{-6} to 10^{-1} M . The parameter k_{cat}/K_m is called the specificity constant because it determines the specificity between competing substrates. At low substrate concentrations k_{cat}/K_m is an apparent second-order rate constant. In order to obtain an accurate determination of k_{cat} and K_m , it is necessary to measure v at $[S]$ where $[S] < K_m$ and at $[S]$ where $[S] > K_m$. In cases where lack of solubility of the substrate makes measurements of v at $[S] > K_m$ impossible, only k_{cat}/K_m can be determined accurately; the individual kinetic parameters k_{cat} and K_m cannot be determined in these cases.

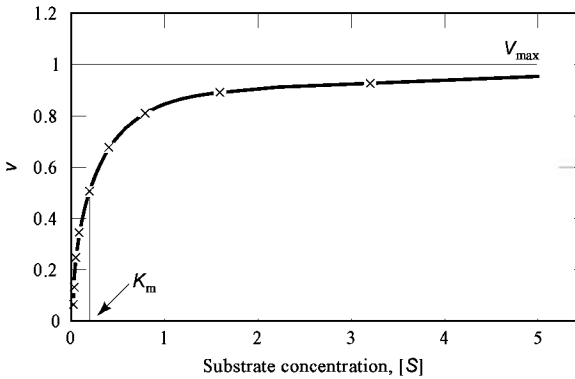


Fig. 1. Reaction velocity v as a function of substrate concentration for a reaction obeying Michaelis-Menten kinetics.

Kinetic parameters are often derived from a set of experimental data by rearranging the data into a linear form, eg, plotting $1/v$ as a function of $1/[S]$ (a Lineweaver-Burk, or double-reciprocal, plot) allows the determination of k_{cat} and K_m from the intercepts of the axes. However, the most accurately determined rates measured at high $[S]$ tend to cluster around the origin, whereas the rate values at low $[S]$, which are least accurately measured, are far from the origin; the latter data tends most strongly to determine the slope. In other words, rearranging the data also rearranges the error distribution, thereby invalidating the assumptions behind linear regression. This method is subject to large errors. However, software using nonlinear regression makes it possible to determine kinetic parameters directly from a plot of v as a function of $[S]$, thus making linear rearrangements unnecessary.

In Figure 2, a double-reciprocal plot is shown; Figure 1 is a nonlinear plot of v as a function of $[S]$. It can be seen how the least accurately measured data at low $[S]$ make the determination of the slope in the double-reciprocal plot difficult. The kinetic parameters obtained in this example by making linear regression on the double-reciprocal data are $v_{\text{max}} = 1.15$ and $K_m = 0.25$ (arbitrary units). The same kinetic parameters obtained by software using nonlinear regression are $v_{\text{max}} = 1.00$ and $K_m = 0.20$ (arbitrary units).

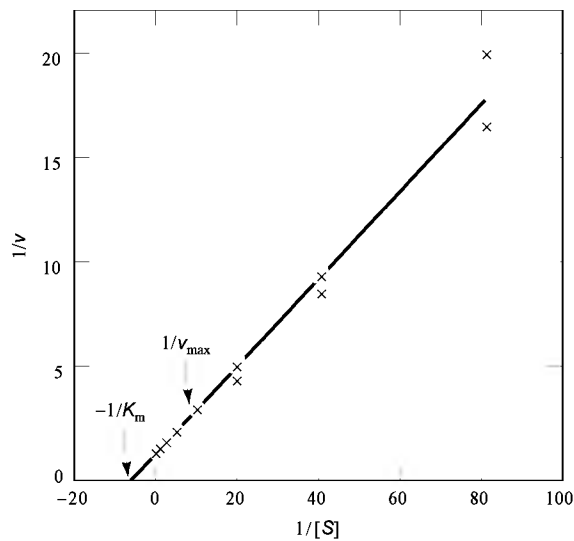


Fig. 2. The double-reciprocal plot of $1/v$ as a function of $1/[S]$. Two determinations of v for each $[S]$ are shown. $1/v_{\max} = 0.87$ and $-1/K_m = -4.05$ (arbitrary units). Figure 1 is a nonlinear regression plot of the same data where $v_{\max} = 1.00$ and $K_m = 0.20$ (arbitrary units).

Enzyme Inhibition. Enzyme inhibitors (qv) are reagents that bind to the enzyme and cause a decrease in the reaction rate. Irreversible inhibitors bind to the enzyme by an irreversible reaction, and consequently cannot dissociate from the enzyme or be removed by dilution or dialysis. Examples of irreversible inhibitors are nerve gases such as diisopropylphosphofluoridate [55-91-4] (DFP).

Reversible inhibition is characterized by an equilibrium between enzyme and inhibitor. Many reversible inhibitors are substrate analogues, and bear a close relationship to the normal substrate. When the inhibitor and the substrate compete for the same site on the enzyme, the inhibition is called competitive inhibition. In addition to the reaction described in equation 1, the competing reaction described in equation 3 proceeds when a competitive inhibitor I is added to the reaction solution.



At high $[S]$, where the number of substrate molecules greatly outnumber the number of inhibitor molecules, the effect of a competitive inhibitor is negligible. Therefore, $v_{\max}$ is unchanged. However, the apparent K_m increases because of a higher degree of inhibition at low $[S]$. The effect of adding a competitive inhibitor is illustrated in Figure 3.

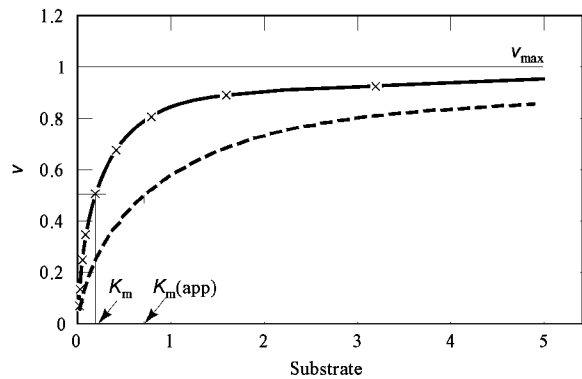


Fig. 3. The effect on kinetic parameters of adding a competitive inhibitor. Reaction velocity v as a function of $[S]$ is shown. ($\times$) Uninhibited reaction; (---) inhibited reaction. As indicated on the figure, the parameter K_m is increased by adding the competitive inhibitor; both curves eventually reach the same $v_{\max}$.

Effect of Temperature and pH. The temperature dependence of enzymes often follows the rule that a 10°C increase in temperature doubles the activity. However, this is only true as long as the enzyme is not deactivated by the thermal denaturation characteristic for enzymes and other proteins. The three-dimensional structure of an enzyme molecule, which is vital for the activity of the molecule, is governed by many forces and interactions such as hydrogen bonding, hydrophobic interactions, and van der Waals forces. At low temperatures the molecule is constrained by these forces; as the temperature increases, the thermal motion of the various regions of the enzyme increases until finally the molecule is no longer able to maintain its structure or its activity. Most enzymes have temperature optima between 40 and 60°C. However, thermostable enzymes exist with optima near 100°C.

The pH dependency of enzyme-catalyzed reactions also exhibits an optimum. The pH optima for enzyme-catalyzed reactions cover a wide range of pH values. For instance, the subtilisins have a broad pH optima in the alkaline range. Other enzymes have a narrow pH optimum. The nature of the pH profile often gives clues to the elucidation of the reaction mechanism of the enzyme-catalyzed reaction. The temperature at which an experiment is performed may affect the pH profile and vice versa.

Enzyme Assays. An enzyme assay determines the amount of enzyme present in sample. However, enzymes are usually not measured on a stoichiometric basis. Enzyme activity is usually determined from a rate assay and expressed in activity units. As mentioned above, a change in temperature, pH, and/or substrate concentration affects the reaction velocity. These parameters must therefore be carefully controlled in order to achieve reproducible results.

Spectrophotometry, a simple and reliable technique, is often used in rate assays. This method can be used when the substrate or the product of the reaction absorbs in the uv or the visible region. In other cases, a nonabsorbing system can be coupled to a system in which the substrate or product absorbs in the uv or visible region.

An example of a direct spectrophotometrical assay is the use of synthetic peptide *p*-nitroanilide substrates to determine protease activity. The *p*-nitroaniline group liberated from the substrates by the protease can be determined spectrophotometrically at 410 nm. An example of an indirect (coupled) spectrophotometric assay is the determination of α -amylase using *p*-nitrophenylmaltoheptaoside. Initially, the substrate is cleaved by the α -amylase and subsequently one of the reaction products, *p*-nitrophenylmaltotrioxide, is cleaved by α -glucosidase, liberating *p*-nitrophenyl, a chromophore

that can be measured at 405 nm.

Potentiometry is another useful method for determining enzyme activity in cases where the reaction liberates or consumes protons. This is the so-called pH-stat method. pH is kept constant by countertitration, and the amount of acid or base required is measured. An example of the use of this method is the determination of lipase activity. The enzyme hydrolyzes triglycerides and the fatty acids formed are neutralized with NaOH. The rate of consumption of NaOH is a measure of the catalytic activity.

Enzyme Nomenclature. The number of enzymes known exceeds two thousand. A system of classification and nomenclature is required to identify them unambiguously. During the nineteenth century, it was the practice to identify enzymes by adding the suffix -in to the name of their source. Names such as papain, ficin, trypsin, pepsin, etc, are still in use. However, this system does not give any indication of the nature of the reaction catalyzed by the enzyme or the type of substrate involved.

By the end of the nineteenth century a more descriptive system was in use. The suffix -ase was appended to the name of the substrate involved in the reaction, eg, amylase, cellulase, protease, lipase, urease, etc. Names that reflected the function of the enzyme with the suffix -ase were also used, eg, invertase, transferase, isomerase, oxidase.

A system based partly on historical names, partly on the substrate, and partly on the type of reaction catalyzed is far from satisfactory. In 1956, the International Union of Biochemistry set up a Commission on Enzymes to consider the classification and nomenclature of enzymes. The Commission presented a report in 1961 whose recommendations for naming and classifying enzymes were subsequently adopted (12). Enzymes are classified on the basis of the reactions they catalyze. Despite its apparent complexities, the system is precise and very descriptive, accommodating existing enzymes and serving as a systematic basis for the naming of new enzymes. All enzymes are placed in one of the six principal classes.

Number	Class	Type of reaction catalyzed
1	oxidoreductases	transfer of electrons
2	transferases	group-transfer reactions
3	hydrolases	transfer of functional groups to water
4	lyases	addition of groups to double bonds or the reverse
5	isomerases	transfer of groups within molecules to yield isomeric forms
6	ligases	formation of C–C, C–S, C–O, and C–N bonds by condensation reactions coupled to ATP cleavage

Each class is divided into groups or subclasses according to the type of reaction catalyzed. These groups are further subdivided according to the nature of the substrate involved. Each enzyme is then assigned a four-digit classification number (EC number) based on this division, and a systematic name to identify the reaction catalyzed. A trivial name may be used if the systematic name is too long or cumbersome. For example, lactase (trivial name) catalyzes the conversion of lactose to galactose and glucose. It is given the systematic name of β -D-galactoside galactohydrolase, and the number *EC* 3.2.1.23, ie, 3-hydrolases; 3.2-glycosidases; and 3.2.1-hydrolyzing O-glycosyl bonds.

Production of Industrial Enzymes

Until about 1950, the predominant method of producing industrial enzymes was by extraction from animal or plant sources; by 1993, this accounts for less than 10%. With the exception of trypsin, chymosin, papain [9001-73-2], and a few others, industrial enzymes are now produced by microorganisms grown in aqueous suspension in large vessels, ie, by fermentation (qv). A small (5%) fraction is obtained by surface culture, ie, solid-state fermentation, of microorganisms (13).

Enzymes are usually sensitive to harsh physical and chemical conditions, and care must be taken during recovery and purification to avoid inactivation of the enzyme. This demands careful selection of production processes and conditions for each individual enzyme. Different methods are subsequently applied to assure the stability and activity of the enzymes during storage and application.

Fermentation. The volume of industrial fermentors range from 20 m³ to several hundred m³, in a few cases exceeding 1000 m³. A conventional fermentor (14) is shown in Figure 4. In most cases, oxygen is required by the microorganism, and air is supplied through a bottom sparger at a rate of 0.5 to 2 tank volumes per minute. The liquid is often agitated to improve gas transfer and mixing. Heat caused by microbial metabolism and agitation is removed through a cooling jacket or coil. Baffles are placed near the wall to increase mixing efficiency and prevent vortex formation. For microorganisms that are very sensitive to shear stress, air lift fermentors (15) can be used. In the air lift fermentor, the rising air bubbles provide the mixing and circulation of the culture broth.

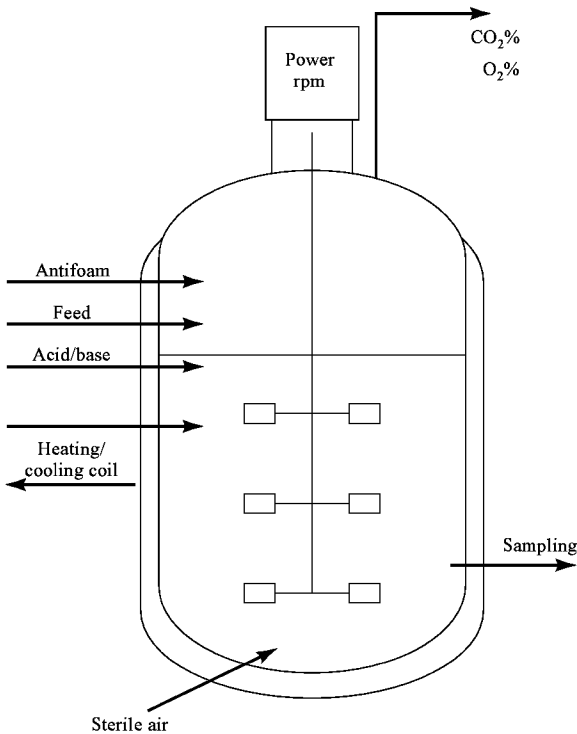


Fig. 4. A conventional fermentor. Foam, total pressure, P_O, pH, temperature, and weight must be monitored and controlled.

To prevent contamination with undesirable microorganisms, the fermentor and auxiliary equipment must be sterilized before inoculation. This is achieved by steam, ie, at least 20 min at 121°C. The incoming air is filtered.

Specific procedures exist for storing (16,17) and propagating microorganisms to obtain reproducible fermentations. The stock culture is stored frozen ($\leq -80^{\circ}\text{C}$) or freeze-dried. To prepare the inoculum (seed) mixture, an aliquot is taken and grown in consecutive solid or liquid cultures of increasing volume. The volume of the last step, the seed fermentor, is typically 4–12% of the main fermentor volume.

The media contains the nutrients for the organism, often agricultural products, eg, starch, soy protein, and palm oil. The initial concentration of raw materials can be as high as 250 kg m⁻³. Various salts, eg, potassium phosphate, are added to provide basic elements. Sometimes vitamins or specific growth factors are also added. A batch fermentation can be divided into a growth phase during which little enzyme is produced and a production phase during which growth is slow. Many enzymes are produced as a result of depriving the microorganism of the carbon and or nitrogen source. Therefore, a slow feed of the limiting nutrient can be introduced to prolong the production period. Such a fed-batch fermentation lasts 3–10 days. In some cases, a continuous fermentation is used where media is constantly added and broth is removed at the same rate. Reported yields of enzymes vary from 1 kg m⁻³ broth to 50 kg m⁻³ (18).

Temperature, pH, and feed rate are often measured and controlled. Dissolved oxygen (DO) can be controlled using aeration, agitation, pressure, and or feed rate. Oxygen consumption and carbon dioxide formation can be measured in the outgoing air to provide insight into the metabolic status of the microorganism. No reliable on-line measurement exists for biomass, substrate, or products. Most optimization is based on empirical methods; simulation of quantitative models may provide more efficient optimization of fermentation.

Recovery. The principal purpose of recovery is to remove nonproteinaceous material from the enzyme preparation. Enzyme yields vary, sometimes exceeding 75%. Most industrial enzymes are secreted by a microorganism, and the first recovery step is often the removal of whole cells and other particulate matter (19) by centrifugation (20) or filtration (21). In the case of cell-bound enzymes, the harvested cells can be used as is or disrupted by physical (eg, bead mills, high pressure homogenizer) and or chemical (eg, solvent, detergent, lysozyme [9001-63-2], or other lytic enzyme) techniques (22). Enzymes can be extracted from disrupted microbial cells, and ground animal (trypsin) or plant (papain) material by dilute salt solutions or aqueous two-phase systems (23).

Ultrafiltration (qv) (uf) is increasingly used to remove water, salts, and other low molecular-weight impurities (21); water may be added to wash out impurities, ie, diafiltration. Ultrafiltration is rarely used to fractionate the proteins because the capacity and yield are too low when significant protein separation is achieved. Various vacuum evaporators are used to remove water to 20–40% dry matter. Spray drying is used if a powdery intermediate product is desired. Lyophilization (freeze-drying) is only used for heat-sensitive and highly priced enzymes.

Sufficient color reduction is often achieved by recovery and purification methods. However, sometimes specific color removal is achieved by adsorption to, eg, activated carbon.

Purification. Enzyme purity, expressed in terms of the percent active enzyme protein of total protein, is primarily achieved by the strain selection and fermentation method. In some cases, however, removal of nonactive protein by purification is necessary. The key purification method is selective precipitation of the product or impurities by addition of salt, eg, sodium sulfate, or solvent, eg, ethanol or acetone; by heat denaturation; or by isoelectric precipitation, ie, pH adjustments. Methods have been introduced to produce crystalline enzyme preparations (24).

Many laboratory techniques have been described to purify proteins (25), but they are often too costly for industrial enzymes, especially column separations. However, aqueous two-phase extraction (26) and ion exchange are used.

Product Requirements. When an enzyme is recovered from fermentation broth, it is usually present in an aqueous solution or processed to a dried state. Both types of preparation have to be formulated to comply with requirements appropriate to their final application.

Requirements related to the storage of enzymes from the time of manufacture to the time of application include enzyme stability, ie, the catalytic activity of the enzyme must remain after prolonged storage at relevant temperatures; microbial stability, ie, industrially recovered enzymes in aqueous solution are potentially excellent growth media for microorganisms, so it is usually necessary to prevent microbial growth; and physical stability, ie, the enzyme must remain in solution to avoid inhomogeneous dosage. Some applications demand special requirements. These can be of a technical nature such as having no precipitate, off-odor, and off-color when mixed into a liquid detergent; and or an absence of side activities, eg, transferase activity must be absent in saccharifying enzymes like amyloglucosidase, and protease must be absent in cell wall-degrading enzymes for the upgrading of vegetable proteins. Other requirements in certain applications derive from approval considerations, eg, only food-grade ingredients, absence of certain microorganisms, and kosher restrictions on enzymes for food applications.

Enzyme Stability. Loss of enzyme-catalytic activity may be caused by physical denaturation, eg, high temperature, drying freezing, etc; or by chemical denaturation, eg, acidic or alkaline hydrolysis, proteolysis, oxidation, denaturants such as surfactants or solvents, etc. pH has a strong influence on enzyme stability, and must be adjusted to a range suitable for the particular enzyme. If the enzyme is not sufficiently stable in aqueous solution, it can be stabilized by certain additives; a comprehensive treatment with additional examples is available (27).

Many enzymes need a certain ionic strength to maintain an optimum stability and solubility, eg, bacterial α -amylases show optimal stability in the presence of 1–2% NaCl. Some enzymes may need certain cations in low amounts for stabilization, eg, Ca²⁺ is known to stabilize subtilisins and many bacterial α -amylases. Antioxidants (qv) such as sodium sulfite can stabilize cysteine-containing enzymes which, like papain, are often easily oxidized.

Polyalcohols, such as glycerol, sugar, sorbitol, and propylene glycol may prevent denaturation (28). Also substrates or substrate analogues often stabilize by conferring an increased rigidity to the enzyme structure.

If a protease is present in solution it is necessary to minimize its activity in order to avoid interference with processing steps, including autolysis during the formulation of a protease. This can be obtained by reducing water activity by means of propylene glycol (25–50%); by adjusting pH to a range where the protease is inactive, eg, low pH in formulations of alkaline proteases; or by adding compounds that reversibly inhibit the protease, eg, formate or borate (29).

Microbial Stability. Microbial growth is hindered by reducing water activity and adding preservatives. An overview is available (30). Reduction in water activity is typically obtained by including approximately 50% of a polyalcohol such as sorbitol or glycerol. Furthermore, 20% of a salt like NaCl has a pronounced growth inhibiting effect.

Some alcohols, eg, propylene glycol, not only lower water activity but also have an additional preservative effect caused by the way they interfere with the cell membrane transport system of the contaminating microorganisms. Surfactants (qv) may show a similar effect.

Organic acids may inhibit growth when present in the undissociated form because of their ability to change the pH inside the cell. The most efficient are benzoic acid and sorbic acid, but formic, acetic, and propionic acid also have this effect. The parabens, ie, *p*-hydroxy benzoic acid esters, are also used because of their antimicrobial effect over a broad pH range.

These systems are allowed in many food applications, but there also exists a range of nonfood preservatives active over a broad pH range. However, these may not be compatible with all enzymes because of their inhibitory or denaturing effects. A useful reference on this subject is available (31).

Physical Stability. Physical stability depends primarily on the purity of the enzyme. Impurities remaining from the fermentation broth may precipitate or form a hazy solution. Unwanted sedimentation is often related to Ca²⁺ or acidic polysaccharides. The solubility of some enzymes can be increased by optimizing the ionic strength or changing the dielectric constant of the solution by adding low molecular-weight polyols.

Formulations. Any formulation is a compromise between the previously mentioned requirements. For example, the fermentation broth may contain enzyme-stabilizing substances, but the application of the enzyme or precipitation problems in the formulation may demand a high degree of purification that eliminates the stabilizers. Alternatively, the pH necessary for good microbial or physical stability may differ from the pH that gives optimum enzyme stability, or a preservative that is effective at the optimum pH for enzyme stability may have a denaturing effect on the enzyme.

The formulation of an enzyme is normally considered a way to store and transport the enzyme until its application. One common exception is immobilized enzymes where formulation is an active part of their application.

It is sometimes possible to add properties in liquid formulations that provide additional functions. Examples in development or in commercial use as of 1993 include microencapsulation (qv) of enzymes for protection against bleach when dispersed in a liquid detergent; addition of certain polymers to protect the enzyme after it has been added to liquid detergents (32), or to boost activity in the final application; addition of surfactants or wetting agents,

eg, α -amylases for textile desizing in order to improve their effect; and addition of buffers to keep the application pH under control.

Immobilization. Enzymes, as individual water-soluble molecules, are generally efficient catalysts. In biological systems they are predominantly intracellular or associated with cell membranes, ie, in a type of immobilized state. This enables them to perform their activity in a specific environment, be stored and protected in stable form, take part in multi-enzyme reactions, acquire cofactors, etc. Unfortunately, this optimization of enzyme use and performance in nature may not be directly transferable to the laboratory.

Cost reduction is the primary argument for using immobilized enzymes, especially when comparing this method with soluble enzyme or nonenzymatic methods; nevertheless, satisfactory technical solutions can be found among the latter two alternatives. When immobilized, expensive enzymes can be re-used, and this may compensate for the cost of immobilization. However, enzyme production methods are constantly being improved and costs reduced accordingly. Costs in this context also include process costs, eg, equipment, energy, product recovery, enzyme inactivation, etc. Enzyme stability factors, eg, temperature, pH, proteases, oxidation, and solvents organics, also are important, but are not often regarded as a cost issue because the desired stability is not always found with soluble enzymes. Continuous processes involving immobilized enzymes enable large substrate volumes to be handled by comparatively small reactors and allow the re-use of enzymes.

A significant advantage of immobilized enzymes is the total absence of catalytic activity in the product. Moreover, the degree of substrate-to-product conversion can be controlled during processing, eg, by adjusting the flow rate through a packed-bed column reactor of immobilized enzyme.

Choice of Method. Numerous enzyme immobilization techniques have been described in the literature; comprehensive books on this and related subjects, including industrial applications, are available (33–36). The more general techniques and some selection criteria are included herein.

Because enzymes can be intracellularly associated with cell membranes, whole microbial cells, viable or nonviable, can be used to exploit the activity of one or more types of enzyme and cofactor regeneration, eg, alcohol production from sugar with yeast cells. Viable cells may be further stabilized by entrapment in aqueous gel beads or attached to the surface of spherical particles. Otherwise cells are usually homogenized and cross-linked with glutaraldehyde [111-30-8] to form an insoluble yet penetrable matrix. This is the method upon which the principal industrial applications of immobilized enzymes is based.

Extracellular microbial enzymes can be immobilized in the form of proteins purified to varying degrees. Cross-linking methods are based on intra- and inter-molecular binding, usually involving the coupling of lysine amino groups by using the bifunctional reagent glutaraldehyde. In many cases this leads to insolubilized, active, and stabilized enzyme. Additional chemicals and proteins may be used in the cross-linking process.

Other immobilization methods are based on chemical and physical binding to solid supports, eg, polysaccharides, polymers, glass, and other chemically and physically stable materials, which are usually modified with functional groups such as amine, carboxy, epoxy, phenyl, or alkane to enable covalent coupling to amino acid side chains on the enzyme surface. These supports may be macroporous, with pore diameters in the range 30–300 nm, to facilitate accommodation of enzyme within a support particle. Ionic and nonionic adsorption to macroporous supports is a gentle, simple, and often efficient method. Use of powdered enzyme, or enzyme precipitated on inert supports, may be adequate for use in nonaqueous media. Entrapment in polysaccharide polymer gels is used for both cells and isolated enzymes.

Membrane reactors, where the enzyme is adsorbed or kept in solution on one side of an ultrafiltration membrane, provides a form of immobilized enzyme and the possibility of product separation.

Microemulsions or reverse micelles are composed of enzyme-containing, surfactant-stabilized aqueous microdroplets in a continuous organic phase. Such systems may be considered as a kind of immobilization in enzymatic synthesis reactions.

The choice of a suitable immobilization method for a given enzyme and application is based on a number of considerations including previous experience, new experiments, enzyme cost and productivity, process demands, chemical and physical stability of the support, approval and safety issues regarding support, and chemicals used. Enzyme characteristics that greatly influence the approach include intra- or extracellular location; size; surface properties, eg, charge pI, lysine content, polarity, and carbohydrate; and active site, eg, amino acids or cofactors. The size, charge, and polarity of the substrate should also be considered.

Industrial-Scale Applications. The breakthrough for immobilized enzyme technology came in the mid-1960s. The first significant development was the production of L-amino acids by continuous optical resolution of synthetic racemates using amino acylase adsorbed onto DEAE-Sephadex (33). The introduction of immobilized penicillin G acylase, and then immobilized glucose isomerase [9055-00-9] (IGI) for the production of high fructose corn syrup (HFCS) demonstrated the enormous industrial potential of immobilized enzymes. IGI is the last in a series of three enzymes used to convert starch to HFCS; ie, liquefaction by α -amylase [9000-90-2], saccharification by glucoamylase [9032-08-0], and isomerization of glucose by IGI to approximately 45° fructose, which is fractionated chromatographically to 55° fructose and 45° glucose. Starch sources other than corn can be used.

The success of IGI was a result of high quality immobilized preparations developed at a time when raw sugar was expensive (1975). When the technology was established and further improved, it remained competitive when raw sugar prices dropped.

A number of factors, concerning both enzyme and process, favored the immobilization of glucose isomerase. The enzyme was expensive because of its low specific activity and its low yield resulting from its intracellular formation, by-product formation could be minimized by using short residence times at enzyme-optimal pH, and particles with physical properties suited to large enzyme columns were needed to handle the huge volumes generated by the starch industry. A commercial product (Sweetzyme) based on glutaraldehyde cross-linked *Bacillus coagulans* cells also was introduced.

Commercial IGIs of the 1990s are based on various enzyme sources, largely *Streptomyces* spp., and include immobilization techniques incorporating the adsorption of purified glucose isomerase (37). An example of an IGI is made by cross-linking *S. murinus* cells with glutaraldehyde and then extruding the resulting matrix into particles. Typical reaction conditions and performance are listed in Table 2.

Table 2. Typical Operating Parameters for Immobilized Glucose Isomerase^a and Penicillin V Acylase^b

Immobilized enzyme	Glucose isomerase	Penicillin V acylase
enzyme reactor	2–5 m height, 0.6–1.5 m diameter	batch + – recycled packed bed
substrate, °	ca 45 w glucose	10 w v penicillin V
initial flow or activity	1.2–6.8 bed vol h	100,000–200,000 units kg
temperature, °C	55–60	35
pH	7–8	7.5 ^c
conversion	42–45° fructose	98° 6-APA
productivity, t kg	15 (dry subst)	0.25
immobilized enzyme at residual act., °	10	50

^a Ref. 37.

^b Ref. 38.

^c Adjusted with NH₃.

The second most important group of immobilized enzymes is still the penicillin G and V acylases. These are used in the pharmaceutical industry to make the intermediate 6-aminopenicillanic acid [551-16-6] (6-APA), which in turn is used to manufacture semisynthetic penicillins, in particular ampicillin [69-53-4] and amoxicillin [26787-78-0]. This is a remarkable example of how a complex chemical synthesis can be replaced with a simple enzymatic one;

four steps, involving toxic and corrosive chemicals, solvents, and a 40°C process step, can be avoided. It is estimated that all 6-APA is now produced by the enzymatic route. Similar reactions on cephalosporins can be made.

Enzyme techniques are primarily developed for commercial reasons, and so information about immobilization and process conditions is usually limited. A commercially available immobilized penicillin V acylase is made by glutaraldehyde cross-linking of a cell homogenate. It can be used in batch stirred tank or recycled packed-bed reactors with typical operating parameters as indicated in Table 2 (38). Further development may lead to the creation of acylases and processes that can also be used for attaching side chains by enzymatic synthesis.

During the 1980s, molecular biology techniques were used successfully to reduce enzyme production costs; this is part of the reason the number of immobilized enzyme processes used in industry are relatively few when compared with soluble enzyme processes, and certainly less than expected in the 1970s. Acylases other than IGI and immobilized penicillin are in use, although they are not of similar significance. Examples from the starch and dairy industry are glucoamylase and lactase. It is expected that the pharmaceutical, organic chemical, and oils and fats industries will increase their use of this technology. Proteases are used in peptide synthesis and nitrilase in the synthesis of acrylamide. Of particular interest is the use of immobilized lipases for optical resolution, ester synthesis, and the production of specific lipids by interesterification (35). These have already demonstrated the feasibility of using immobilized enzyme technology in nonaqueous or low water systems.

Granulation of Enzymes. Although the trend is to market industrial enzymes as liquid products, solid enzyme is needed. Examples of this are enzymes for solid detergents, animal feed, and flour improvement.

Several different methods have been used for the granulation of enzymes. In general, the development of granulation methods focuses on parameters such as the cost of the process and solubility of the granulate. The focus for detergent enzymes is directed onto a single parameter, ie, a dust-free product. It should not contain particles that can become airborne, and the particles must be strong and resilient in order to avoid creating dust during handling of the enzyme in the detergent factory (39).

To reduce the enzyme dust level in detergent factories to an absolute minimum, the majority of detergent enzyme granulates are coated with a layer of inert material. This coating can also be used for coloring purposes. The color of the enzyme granulate itself is often brown. Titanium dioxide [13463-67-7] can be added to the coating medium to make the product whiter; alternatively, another attractive color can be added.

For granulation of enzymes for purposes other than incorporating into detergents, the same methods can be used as for the detergent enzymes, although some of the additives needed for the production of a rigid granulate may not be accepted, ie, in certain enzymes for the food industry. Fortunately for such applications the handling of the enzyme is more gentle, and the requirement for physical stability less. For these enzymes, a fluid-bed granulation performed as an agglomeration of powder with a liquid binder, or as a coating of the enzyme onto inert carrier particles of selected size, often gives the desired quality of product. These enzyme preparations may also be coated.

Various methods are used for evaluating the quality, ie, physical strength and enzyme dust formation, of the granulate. In the elutriation process, a sample of product is fluidized in a glass tube with a perforated bottom plate for 40 minutes. Dust from the sample is collected on a filter and the enzyme activity measured. An acceptable dust level is when less than 5–10 ppm of the activity of the sample has been collected. In the so-called Heubach method, 20 g of granulate is elutriated. During the elutriation, four steel balls are rotated in the bed in order to evaluate the impact of attrition on the dust release of the enzyme. The dust is collected on a filter and measured. The acceptable dust level is very low.

Solid Preparation Methods.

Prilling. A suspension of the enzyme is spray cooled in waxy material, eg, an ethoxylated fatty alcohol, with a melting point over or about 50°C, to form small beads. Bead size may be varied and surface coatings may be applied.

Extrusion. The enzyme is mixed with binders, diluting agents such as inert salts, and the appropriate amount of water. The resulting thick paste is then extruded to small rods having a diameter of about 0.6 mm. Preferably the extrudate is then further spheronized to a product with a more rounded form.

Granulation. The enzyme is diluted with inert material and binders, as described above, but with the option of adding fibers which act to reinforce the granules. It is then granulated in a medium or high shear granulator.

Industrial Applications

Detergent Enzymes. Detergents, from Latin *detergere*, to clean, is here understood to mean cleaning products in a broad sense, ie, not only products for household laundering, including soaking and topical spot removal, but also automatic dishwashing detergents (ADDs) and products for a wide range of industrial and institutional (I&I) cleaning functions.

Penetration of enzymes in laundry detergent products is close to 95% in Europe, Japan, and the United States. Enzymes have become an important ingredient in detergents, along with surfactants, bleaching systems, and builders (see DETERGENCY). Several changes in washing habits and detergent formulations underlie this development.

Lower washing temperatures, required by modern textiles made from synthetic fibers and for reasons of energy conservation, have caused a demand for new, more effective, detergent ingredients. Compacted laundry detergent products compensate for reduced levels of other detergent ingredients by increasing enzyme dosages, resulting in concentrations of up to several mg enzyme protein per liter of washing liquor, as opposed to usually not more than 1 mg/L for non-compact detergents. ADDs with high alkalinity and chlorine bleach are replaced by less aggressive and environmentally more friendly detergent products containing an active oxygen bleach system plus an enzyme system. Finally, the detergent enzyme producers are able to offer products with a more favorable price/performance ratio than previously.

Laundry Soilings. Important classes of soilings to be removed by laundry detergents include water-soluble soilings, eg, soluble inorganic salts, sugar, urea, and perspiration; proteinaceous soilings, eg, blood, egg, milk, cutaneous scales, grass, and spinach; starch; fats/oils, eg, animal fats, vegetable fats, sebum, waxes, and mineral oil; particulate matter, eg, insoluble metal carbonates, oxides, silicates (clay), carbon black, dust, and humus; and bleachable stains, eg, fruit and vegetable juices, wine, tea, and grass. Stains with good water solubility are easily removed during the washing process. All other stains are partially removed by the surfactant-builder system of a detergent, although the result is often unsatisfactory, depending on the conditions. In most cases a suitable detergent enzyme may help. Contrary to the purely physical action of the surfactant system, enzymes work by degrading the dirt into smaller and more soluble fragments. However, to remove a stain totally requires the joint effects of the enzyme, surfactant system, and mechanical agitation.

Protein and starch stains are removed by proteases and amylases, respectively. Fats and oils are generally difficult to remove at low wash temperatures by conventional detergents. By using lipases, it is possible to improve the removal of fats/oils of animal and vegetable origin even at temperatures where the fatty material is in a solid form. Particulate soils can be difficult to remove, especially if the particle size is small. Removal of particulate soil from cotton fabric can be improved by use of a cellulase which removes cellulose fibrils from the surface of the yarn.

Various kinds of dirt may adhere to textile surfaces via a glue of proteinaceous, starchy, or fatty material. In such cases of anchored dirt, an enzyme may assist in removing the dirt even though it does not attack the dirt directly. Compounds from several of these classes are intimately mixed in combined soilings, eg, human sebum on shirt collars, cocoa milk, gravy, or chocolate. Bleachable stains are the only group of stains for which no enzyme product has been marketed (ca 1993). However, the patent literature indicates that efforts are being made to develop an enzymatic bleaching system.

Laundering Conditions Around the World. Any laundering process is an interplay between the equipment used; the materials entering the process, ie, detergent, additional bleach, fabric softener, or water; and the procedure followed. Equipment and procedures in three principal geographical areas are summarized in Table 3.

Table 3. Washing Equipment and Procedures, Worldwide^a

Conditions	U.S./Canada ^b	Japan	Western Europe
machine type ^c	agitator	impeller	drum ^d

fabric load kg	2–3	1–1.5	3–4
wash liquor L	35–80	30–45	18–25
wash temperature, °C	50 (hot)		
	27–43 (warm)	10–40	90
	10–27 (cold)		60
			40
			30
water hardness CaCO ₃ , ppm	relatively low 100	very low 50	relatively high 250
recommended detergent dosage ^e , g L	1–5	1–3	8–10

^a Information mainly extracted from Ref. 40.
^b Chlorine bleach commonly added to wash.
^c Drying process by automatic dryer.
^d Heating coils used.
^e In the United States and Japan usually without bleaching components.

Detergent compositions also vary from country to country. The world market for household detergents can be divided into four segments according to the physicochemical properties of the wash solutions prepared from the detergents. Low pH, low ionic strength detergents are liquid detergents having pH from 7.5 to 9. They contain no bleach and only low levels of salts. Medium pH, medium ionic strength detergents are typically compact powder detergents from Japan and regular powder detergents from the United States. Their pH is about 9 and they contain no bleach. High pH, high ionic strength detergents with bleach have a pH from 9.5 to 10.5, and contain an activated bleaching systems, eg, European regular powder detergents. They also contain sodium sulfate as a filler, and builder systems, eg, sodium triphosphate, zeolite, and sodium carbonate. High dosages are used, which give a wash liquor with a high ionic strength. High pH, medium ionic strength detergents with bleach is a new group of detergents, eg, European compact powder detergents which have gained a huge market share during the early 1990s. They are compact as a result of removal of most or all of the sodium sulfate; pH is from 9.5 to 10.5.

Detergent Enzymes Performance Evaluation. Not all enzymes with a potential for stain degradation and or removal are suitable for inclusion in detergent products. A detergent enzyme must have good activity at the pH of detergent solutions (between 7 and 11) and at the relevant wash temperatures (20 to 60°C), and must be compatible with detergent components during storage as well as during the wash process, eg, surfactants, builders, bleaches, and other enzymes. In particular, such an enzyme must be resistant toward protease degradation under these conditions. With enzymes like proteases and lipases, for which the average load of dirty laundry contains a multitude of different substrates, a broad substrate specificity is demanded. A given enzyme may be assayed by its action on soluble substrates under chemical and physical conditions different from those encountered in a real-life wash. Such experiments indicate the enzyme's performance with respect to pH and temperature variations, or in conjunction with other soluble substances, etc. The analytical data thus obtained are not necessarily representative of the wash performance of the enzyme, and real wash trials are necessary to evaluate wash performance of detergent enzymes.

Wash trials are carried out by the use of soiled test pieces, eg, commonly used stains for protease evaluation are milk, blood, and grass. Commercial pre-soiled test pieces also may contain particulate matter, eg, carbon black, as part of the stain matrix. Test materials are available ready-to-use from a number of research and testing institutes in Europe and the United States, eg, Center for Testmaterials, Vlaardingen, Holland; Wascherei Forschungs Institut, Krefeld, Germany; EMPA, St. Gallen, Switzerland; Instituut voor Reinigingstechnieken TNO, Delft, Holland; and Testfabrics, Middlesex, New Jersey. Alternatively, enzyme manufacturers can supply preparation procedures.

Laboratory wash trials are usually conducted in small-scale models of washing machines. The Terg-o-tometer simulates the top-loaded U.S. type of washing machine, and the Launder-ometer simulates the European drum-type machine (41). The evaluation of the effects on the test pieces can be made visually or by measuring the reflectance of light under specified conditions. Typically, the intensity of light remitted at 460 nm when illuminating the test pieces with a standardized daylight source is expressed as a percentage, R, of the intensity of incident light at the same wavelength. The ΔR value is then a measure of the enzyme effect; it is defined as the difference in R between fabric washed with and without enzyme. The R value is known to correlate well with the visual impression of whiteness of the fabric, and depending on type and degree of soiling, differences in R of 2–3 units are recognizable to the human eye.

Figure 5 records protease performance as a function of enzyme dosage. The performance is better at 60°C than at 40°C until a dosage of approximately 0.05% (w/w) of the enzyme granulate in the detergent is reached. At this enzyme level, the fabric in both cases is clean, ie, the measured R value is identical to that of an unsoiled test piece washed under the same conditions.

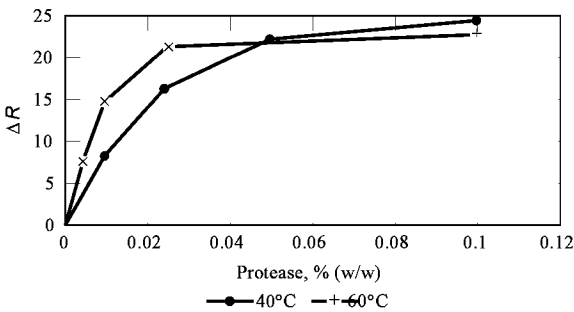


Fig. 5. Protease performance of Savinase 6.0 T (Novo) in powder detergent, 4 g L in a Launder-ometer test. (●) 40°C for 47 min; (×) 60°C for 40 min.

Detergent enzyme performance is often reported in the form of such dose-response curves. The performance increases dramatically at the beginning, but reaches a maximum level at higher enzyme concentrations. The extent to which the enzyme is able to remove stains from the fabric depends on the detergent system, temperature, pH, washing time, wash load, etc. Enzyme wash performance varies between liquid and powder detergents and with the composition of the soiling (Fig. 6).

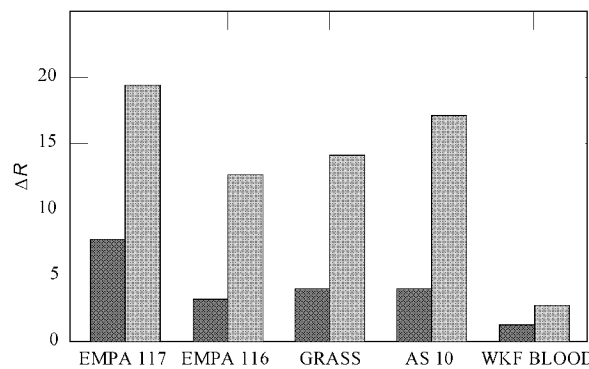


Fig. 6. Washing performance on different soilings of a U.S. liquid detergent (■) and a U.S. powder detergent (▨) in a Terg-o-tometer operating at 20°C for 10 min; one enzyme dosage. EMPA 117 (milk, blood, and ink on polyester cotton); EMPA 116 (milk, blood, and ink on 100° cotton); grass on (100° cotton); AS 10 (milk, oil, and pigments on 100° cotton); blood soilings (on 100° cotton).

Enzyme Types.

Proteases. A protease is an enzyme that hydrolyzes proteins into smaller fragments, ie, peptides or amino acids. A detergent protease degrades the insoluble protein in stains into fragments that can be removed or dissolved by the detergent. All detergent proteases of the early 1990s are produced by *Bacillus* strains. Some are highly alkaline, ie, have a maximum activity in the high pH range, such as Maxacal and Savinase. Others are low alkaline, eg, Maxatase, Alcalase, and Subtilisin Novo (BPN'). They all have a molecular weight between 20,000 and 30,000, and a serine in the active site of the enzyme.

Protease performance is strongly influenced by detergent pH and ionic strength. Surfactants influence both protease performance and stability in the wash solution. In general, anionic surfactants are more aggressive than amphoteric surfactants, which again are more aggressive than nonionic surfactants.

All detergent proteases are destabilized by linear alkylbenzenesulfonate (LAS), the most common type of anionic surfactant in detergents. The higher the LAS concentration and wash temperature, the greater the inactivation of the enzyme. The presence of nonionic surfactants, however, counteracts the negative effect of LAS. Almost all detergents contain some nonionic surfactant; therefore, the stability of proteases in a washing context is not problematic.

Builders, eg, sodium triphosphate and nonphosphate builders such as zeolite and citrate, remove free calcium from the washing solution. Co-builders such as nitrilotriacetic acid or polycarboxylates also may be incorporated into the detergent formulation. Wash performance of detergents decreases with increasing calcium concentration. Protease performance varies, but high calcium concentrations tend to reduce protease performance. Therefore it is an advantage to add a builder system to the detergent. Proteases need a small amount of calcium for the sake of stability, but even with the most efficient builder systems, stability during wash is not a problem.

Bleach systems oxidize proteinaceous stains on fabric, often making the stains more difficult to remove. Detergent proteases can counteract this negative effect of the bleach system. The most commonly used bleach systems in detergents consist of sodium perborate [7632-04-4] plus an activator such as tetraacetythylenediamine [10543-57-4] (TAED). Perborate releases hydrogen peroxide [7722-84-1], H₂O₂, which combines with the activator to form a peroxycarboxylic acid. Most detergent proteases are stable during the wash cycle in the presence of such active-oxygen bleach systems. However, storage stability in detergents containing bleach may be a problem with the established detergent proteases. New protein-engineered proteases introduced onto the market replace the most bleach-sensitive amino acid, ie, methionine close to the active site, with other amino acids not sensitive toward oxidation. This slight change in the molecular structure significantly increases the storage stability in detergents containing bleach (42) (Fig. 7). Chlorine bleach [7681-52-9] (sodium hypochlorite), NaOCl, is not incorporated into laundry detergents themselves, but is used separately in some parts of the world as an additive. In normal wash concentrations above 200 ppm it quickly oxidizes the enzymes, resulting in loss of protease activity.

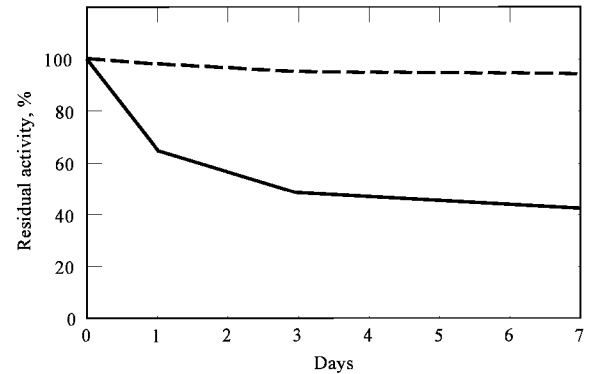


Fig. 7. Storage stability of proteases in European powder detergent with activated bleach system. (—) Traditional protease; (---) protein-engineered protease.

Most ingredients in a detergent formulation contribute to the ionic strength of the wash solution. The effect of ionic strength on protease performance depends on pH and enzyme identity. The pH wash solutions also affects protease performance (Fig. 8).

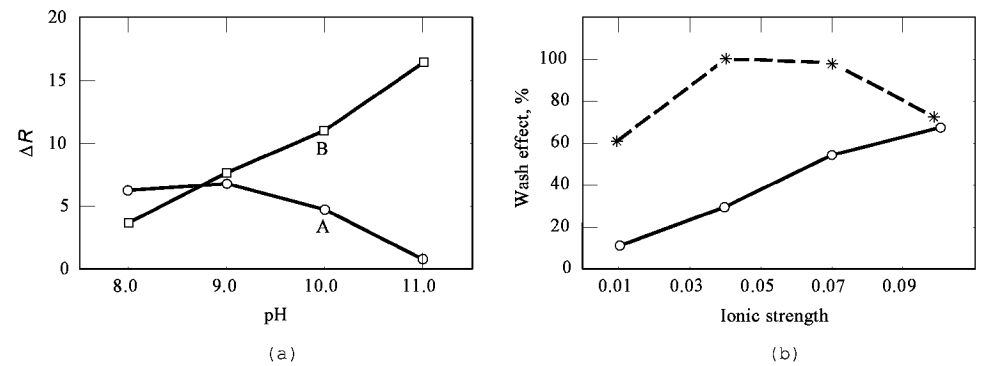


Fig. 8. Protease washing performance in a U.S. liquid detergent. Grass soiling in a 10 min wash at 30°C with one enzyme dosage. **(a)** pH profile of commercial proteases A and B. **(b)** Effect of increasing ionic strength, adjusted with Na₂SO₄, of commercial protease B at (—) pH 8 and (---) pH 11.

Amylases. Commercial laundry amylases comprise the α -amylase from *Bacillus amyloliquefaciens* and the heat-stable α -amylase from *Bacillus licheniformis*. Alpha-amylases are characterized by attacking the starch polymer in an endo fashion, randomly cleaving internal 1,4-bonds to yield shorter, water-soluble dextrans. They are the preferred type of amylase for laundry detergents, and are included in both powder and liquid formulations in many countries. Alpha-amylases boost overall detergent performance at lower wash temperatures and with milder detergent chemical systems. They catalyze the degradation of starch stains, and improve cleaning by hydrolyzing the starch glue that binds other dirt and stains to fabric. A noticeable amylase effect is obtained in the main wash, as well as with prespotting, ie, dramatically increased enzyme concentration, and presoaking, ie, prolonged reaction time.

Examples of artificially soiled test pieces used to test the performance of amylases include cocoa milk sugar, cocoa sugar potato starch, and starch carbon black, all on cotton or polyester cotton.

Bacterial α -amylases used in laundry detergents are fully compatible with detergent proteases, ie, the two enzymes work together in the wash process. During storage in both powder and liquid detergents, the amylases are very stable in the presence of proteases.

Lipases. The idea of using lipases in the wash process dates back to 1913 when O. Rühm suggested adding pancreatin [8049-47-6] to detergent formulations. Many patents have demonstrated that lipases can improve the removal of fatty stains when used in powder and liquid detergents, special presoakers, or other cleaning agents. Intense research activity is also reflected in the literature (43–45).

Only one detergent lipase, ie, Lipolase introduced by Novo in 1988, has been marketed. The first household powder detergent containing lipase was introduced in Japan in the same year; in Europe and the United States in 1990–1991. Lipase is often incorporated into the new compact powder formulations.

The slow development of a commercial detergent lipase was due to low fermentation yields and to difficulties in finding lipases with the appropriate characteristics for application in household detergent products. Patent literature indicates the only feasible way to produce lipases at an acceptable cost performance ratio is by genetic engineering (46–48). For example, Lipolase, mol wt 32,000, was originally isolated from the fungus *Humicola lanuginosa* with low levels of enzyme expression; using rDNA techniques the lipase is expressed in acceptable yields in the harmless host microorganism *Aspergillus oryzae* (46,49).

The natural substrates for lipases are triglycerides, the main constituents of stains originating from animal fat and vegetable oil. A triglyceride molecule is composed of three fatty acid moieties linked to a glycerol backbone by ester bonds. Lipases catalyze the hydrolysis of the ester bonds, giving a mix of free fatty acids, diglycerides, monoglycerides, and glycerol. The more degraded the fatty stain becomes, the easier it is to remove from the fabric due to its increased hydrophilicity.

Because of the presence of free fatty acids in the mix of hydrolysis products, pH strongly influences the removal of decomposed stains. The best rate of removal requires a pH value above 8 (50).

Lipases are also active during a certain period of the drying step (51), eg, Lipolase displays maximum activity when the moisture content on the fabric is 20–30% by weight. This means that significant decomposition of any residual fatty matter will take place while the laundry is drying. This hydrolytic activity does not result in an immediate advantage in terms of fat removal; however, next time the stained fabric is washed the stain will be removed more effectively.

Lipases have proven to be effective in prespotters and other liquid detergent formulations when used in undiluted form for pretreatment of tough fatty stains. The low water content on the fabric in this situation is believed to be responsible for the high catalytic activity (50).

Cellulases. A 1970 patent application (52) covered the idea of using cellulases to prevent fabrics from becoming harsh as a result of repeated washings (see TEXTILES, FINISHING). Low yields and expensive production methods hindered the market development of cellulases. Very few cellulases that are stable and active at high pH were identified before the late 1980s. Celluzyme from Novo Nordisk and Bio-Beeds from Kao, originating from fungal and bacterial species, respectively, are the only alkaline detergent cellulases available in the early 1990s.

Cellulases cleave β -1,4-glycosidic bonds in cellulose and work on natural textile fibers such as cotton and blends containing cotton. Cleaning by removal of particulate soils, softening, and improved color brightness are the three basic benefits obtained from cellulases. When a textile is exposed to shear stress, either during wear or washing and tumble-drying, the surface becomes slightly damaged. A close look at the yarn reveals the presence of fibers and fibrils ranging in size from a few μ m to a few mm on the surface. The damaged textile surface scatters light, giving the fabric a grayish or dull appearance, and affecting color brightness. Dust particles also tend to stick to the damaged areas, adding to the gray appearance. If the material features a number of colors, the contrast between adjacent areas of different colors will be reduced. Damaged fibers are also thought to be responsible for making the fabric surface more rigid, thereby reducing softness. Cellulases hydrolyze exposed β -1,4 bonds in the cellulose, which leads to removal of the fibrils. This is believed to be the mechanism behind the softening and color-brightening effects. Published literature indicates that little is known about the mechanism behind the cleaning action of cellulases. Possible explanations for the cleaning effect are that by removing the fibrils, the soil attached to them is released, and that the enzymatic action facilitates cleaning by exposing dirt trapped in the fiber matrix to the washing solution (53,54).

Softening effects, similar to those obtained with cellulases, can also be achieved by using claylike compounds in the detergent or cationic surfactants in fabric softening additives used in the rinse. These materials are thought to achieve their effect by coating the fibers with a lubricating layer, thereby lowering the friction and reducing the tendency of the fibrils to bond together (52). The use of cationic surfactants is not restricted to cotton or mixed cotton fabrics. However, coating fibrils by cations may lower the water absorption properties of the fabric. In some cases, eg, towels, this is undesirable and cellulases are considered the superior softening agents. Cellulases also may be preferable from an environmental point of view.

There is concern that enzyme activity of cellulases may cause loss in textile strength and or weight loss. Removal of fibers fibrils lowers the weight to some extent, but since the damaged fibers contribute only little to textile strength it should be possible to have negligible loss of strength by choosing a suitable enzyme concentration. Damage to the fabric itself should only occur at higher concentrations. Appropriate enzyme levels must be determined for any specific formulated product under the intended conditions of use.

Industrial and Institutional Cleaning. The application of enzymes has grown substantially within the industrial and institutional (I&I) sector. The primary field is the use of detergent enzymes for laundry purposes, but a broad range of other applications has been investigated, eg, I&I dishwashing, membrane cleaning, drain and bowl cleaning, cleaning of septic tanks and sewage plants, hard surface cleaning of walls or machinery parts where a cleaning-in-place procedure can be used, and cleaning of apparatus parts like endoscopes and electrodes. The choice of the relevant enzyme type is directly related to the composition of the soils, waste, or deposit that has to be dissolved and removed. Thus, as for household laundering, proteolytic, amylolytic, cellulytic, or lipolytic enzymes can be considered. Proteases are suitable for use on fabrics heavily soiled with blood and or meat residues, eg, from hospitals and the food industry, in particular slaughterhouses; fatty stains are removed efficiently from restaurant tablecloths and napkins by the addition of a lipase to the detergent; residues of starchy foods such as mashed potatoes, spaghetti, hot oatmeal, and chocolate are cleaned with the use of amylases.

The I&I cleaning procedures as a whole, compared with household laundering, are characterized by huge variations in the composition of the soils, types of surface to which they adhere, cleaning time available, etc. The optimum choice of enzyme type and dosage level normally has to be established through a cooperation between the customer (end user), manufacturer of the detergent, and enzyme producer.

Automatic Dishwashing. There are many differences between laundering and automatic dishwashing. The hard surfaces present in the latter process differ from textiles because they are impermeable to soils; therefore, cleaning fluids have better access to the soils.

During the 1970s, manufacturers tried to add enzymes to automatic dishwashing detergents (ADDs), which required that they leave out chlorine bleach, ie, hypochlorite precursors. Because there was no adequate substitute for chlorine bleach, tea and coffee stains were left behind. In the early 1990s the availability of active-oxygen bleaches, compatible with enzymes, and increasing concern about safety and environmental issues, are causing extensive changes in ADD formulations, especially in Europe.

Earlier formulations contained mainly chlorine bleach, metasilicates, triphosphate, and nonionic surfactants. Modern manufacturers have switched to more complicated formulations with disilicates, phosphates or citrate, phosphonates, polycarboxylates, nonionic surfactants, oxygen bleach, bleach activator, and enzymes. The replacement of metasilicates by disilicates lowers pH from approximately 12 to 10.5, at 1 g ADD L water. The combined effect of decreased pH, the absence of hypochlorite, and the trend toward lower wash temperatures has paved the way for the introduction of enzymes into ADDs. Most ADD brands in Europe are part of the new generation of ADD products with enzymes. The new formulations are described in the patent literature (55–57).

ADD enzymes used are heat-stable proteases and α -amylases. The actions of the enzymes in ADDs are similar to those in laundering. The performance of enzymes in automatic dishwashers can be evaluated on selected stains (58) or according to standard procedures described for the testing of dishwashers involving a series of standard soils (59,60). However, some soils in the standard procedures, such as food residues cross-linked by heat, are difficult to remove (61) and are not obvious substrates for enzymes. Starch soils are considered the most stubborn kind of soil on kitchenware. This applies to freshly formed deposits as well as to the starchy film that tends to build up on plates leaving them with a dull appearance. Several α -amylases exist with a high temperature optimum, approximately 70°C (62), that are efficient at removing starch film even under the harsh conditions existing in a dishwasher (63).

Protein residues, eg, soft-boiled egg yolk, are difficult stains to handle. If the stains are not totally denatured, proteases can decompose them. There are commercial proteases with a high temperature optimum (60°C) that can remove most protein soils in a dishwasher (63). Patents on the use of lipases in ADDs have claimed that lipases can reduce the formation of spots and films on glasses (62,64–66); however, no commercial application of lipases in ADDs has been implemented.

Starch Conversion. In the early 1960s the first amyloglucosidase type of enzyme was launched. This enabled starch to be broken down completely into glucose. Shortly afterward, almost all glucose production was reorganized for enzymatic hydrolysis instead of acid hydrolysis; advantages included greater yields, higher degree of purity, and facilitated crystallization. In 1973 an immobilized glucose isomerase was developed that made the industrial production of fructose syrup feasible. In the 1990s the starch processing industry is the second largest consumer of enzymes. Because of the many different products made from starch, significant research efforts have been devoted to this area of application.

The primary steps in the conversion of starch are liquefaction, saccharification, and isomerization. By controlling the enzymatic reactions, sugars of different sweetness can be produced to suit the various needs of manufacturers of food and nonalcoholic beverages.

A slurry of the starch is cooked in the presence of a heat-stable bacterial endo- α -amylase. The enzyme hydrolyzes the α -1,4-glycosidic bonds in pregelatinized starch, the viscosity of the gel rapidly decreases, and maltodextrins are produced. The process may be terminated at this point. The solution is purified and dried, and the maltodextrins are utilized as blandtasting functional ingredients in dry soup mixes, infant foods, sauces and gravy mixes, etc.

Further hydrolysis using an amyloglucosidase leads to the formation of sweet-tasting, fermentable sugars. Sweet starch hydrolyzates with special functional properties may be obtained by using a fungal α -amylase either on its own or in combination with an amyloglucosidase. Alternatively a vegetable β -amylase can be used. The sweet syrups and dextrose are used in beverages, confectionery, canned fruit, bakery products, and ice cream.

Many products made by fermentation are also based on the conversion of starch. Some examples of the use of enzymatically hydrolyzed starches are the production of alcohol, ascorbic acid, enzymes, lysine, and penicillin.

Starch Liquefaction. Starch in its natural state is only degraded slowly by α -amylases. To make the starch susceptible to enzymatic breakdown, it is necessary to gelatinize and liquefy a slurry with a 30–40% dry matter content. Gelatinization temperature depends on the type of starch (67); corn is the most common source of industrial starches followed by wheat, tapioca, and potatoes. Liquefaction is achieved by adding a heat-stable α -amylase to the starch slurry. The equipment used for liquefaction may be stirred tank reactors, continuous stirred tank reactors (CSTR), or a jet cooker. Most starch processing plants liquefy the starch with a single enzyme dose in a process using a jet cooker (Fig. 9).

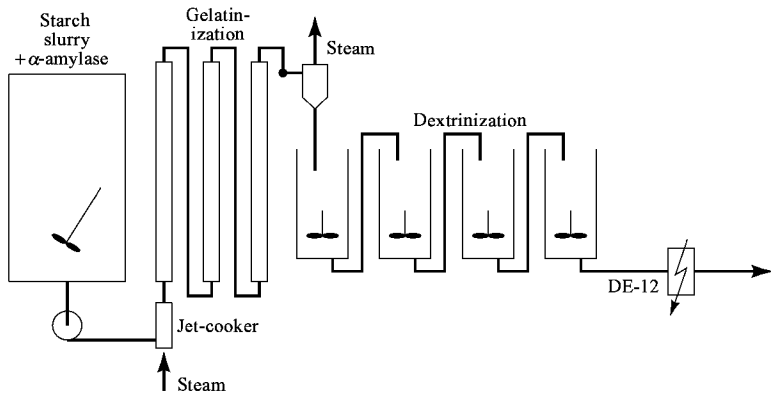


Fig. 9. Starch liquefaction process, 35% dry substance at pH 6.3, 40 ppm Ca^{2+} . Single enzyme dose of 0.5 kg α -amylase 120 L starch slurry, at 105°C for 5 min (gelatinization), followed by 95°C for 2 hours (dextrinization) (68).

In the alcohol industry, grain or potato raw materials are milled and water added to form a slurry or mash which is heated either batchwise or continuously. Traditionally, the mash is heated to 150°C by the injection of live steam. To reduce viscosity, α -amylases are added both during heating to 150°C and during cooling. Thermostable α -amylases from *Bacillus licheniformis* are the most commonly used enzymes for these processes (68).

In German batch processes, raw materials are not milled; gelatinization is achieved by cooking with live steam in a Henze cooker. No addition of enzyme or mechanical agitation is necessary during the cooking stage. The cooked mash is blown through a counter pressure valve into the mash tub where liquefaction takes place either at high (80°C) or low (55–60°C) temperature.

Cooking extruders have been studied for the liquefaction of starch, but the high temperature inactivation of the enzymes in the extruder demands doses 5–10 times higher than under conditions in a jet cooker (69). For example, continuous nonpressure cooking of wheat for the production of ethanol is carried out at 85°C in two continuous stirred tank reactors (CSTR) connected in series; plug-flow tube reactors may be included if only one CSTR is used (70).

During liquefaction, the α -1,4 linkages are hydrolyzed at random. This reduces the viscosity of the gelatinized starch, and increases the dextrose equivalent (DE), ie, a measure of the degree of hydrolysis of the starch. The liquefaction is carried out in such a way as to give the required DE for the subsequent process. For saccharification to dextrose, a DE of 8–12 is normal. Higher DE values are often necessary in maltodextrin production. The maximum DE obtainable is approximately 40. The important α -amylases used in industry originate from *Bacillus licheniformis*, *Bacillus subtilis*, and *Aspergillus oryzae*. Table 4 lists application conditions and key characteristics of enzymes available for starch processing.

Table 4. Starch-Degrading Enzymes of Industrial Importance

Enzyme	Origin	Application		Ca^{2+} , ppm ^a
		Temp., °C	pH	

α -amylase				
bacterial, mesophilic	<i>Bacillus subtilis</i>	80–85	6–7	150
bacterial, thermophilic	<i>Bacillus licheniformis</i>	95–105	6–7	20
fungal	<i>Aspergillus oryzae</i>	55–70	4–5	50
pullulanase	<i>Bacillus acidopullulyticus</i>	55–65	3.5–5	0
amyloglucosidase	<i>Aspergillus niger</i>	55–65	3.5–5	0

^a Minimum dosage.

β -Amylases are exoenzymes that attack amylose chains and result in the successive removal of maltose units from the nonreducing end. In the case of amylopectin, the cleaving stops two to three glucose units from the α -1,6-branching points. β -Amylase [9000-91-3] is used for the production of maltose syrups and for adjunct processing in breweries. The most important commercial products are made from barley or soybeans.

Isoamylase [9067-73-6] (glycogen-6-glucanohydrolase) and pullulanase [9012-47-9] (pullulan-6-glucanohydrolase) hydrolyzes α -1,6-glucosidic bonds of starch. When amylopectin is treated with a pullulanase, linear amylose fragments are obtained. Using a heat- and acid-stable pullulanase in combination with saccharifying enzymes makes the starch conversion reactions more efficient (71).

Malted barley contains α - and β -amylases along with proteases and phytases. Most standardized microbial enzyme preparations for industrial starch conversion contain approximately 100 times more amylase activity than malt. In beermaking, malt is not just valuable for its enzymes but also for flavor compounds.

Saccharification of Liquefied Starch. Maltodextrin [9050-36-6], DE of 15–25, produced from liquefied starch is commercially valuable for its rheological properties. Maltodextrins are used in the food industry as fillers, stabilizers, thickeners, pastes, and glues. When maltodextrins are saccharified by further hydrolysis using an amyloglucosidase or fungal α -amylase, a variety of sweeteners can be produced with DE values of 40–45 (maltose), 50–55 (high maltose syrups), and 55–70 (high conversion syrup) (72). By applying a series of enzymes including β -amylases, glucoamylases, and pullulanases for debranching, it is possible to produce conversion syrups with maltose contents close to 80° (71). A syrup with a glucose content of 95–97° can be produced from most starch raw materials, eg, corn, wheat, potatoes, tapioca, barley, and rice.

An obstacle in the saccharification of starch has been the α -1,6 bonds, ie, the branch point barriers. Amyloglucosidases, introduced in the early 1960s, hydrolyze the α -1,4 bonds rapidly, but the α -1,6 bonds are hydrolyzed much more slowly. By using a pullulanase in conjunction with the amyloglucosidase (AMG) at the start of the saccharification, the α -1,6 bonds of the branched dextrins can also be hydrolyzed rapidly. As a result, fewer branched oligosaccharides accumulate toward the end of the saccharification. This reduces the tendency for glucose molecules to condense to mostly isomaltose. The point at which the reverse reaction outweighs glucose formation is thus shifted toward a higher glucose concentration (Fig. 10). Figure 11 indicates that AMG and pullulanase together yield higher glucose than AMG alone. It also illustrates the influence of the dry substance (DS) on the maximum yield of glucose. Carrying out the saccharification at lower dry substance levels results in higher glucose yields.

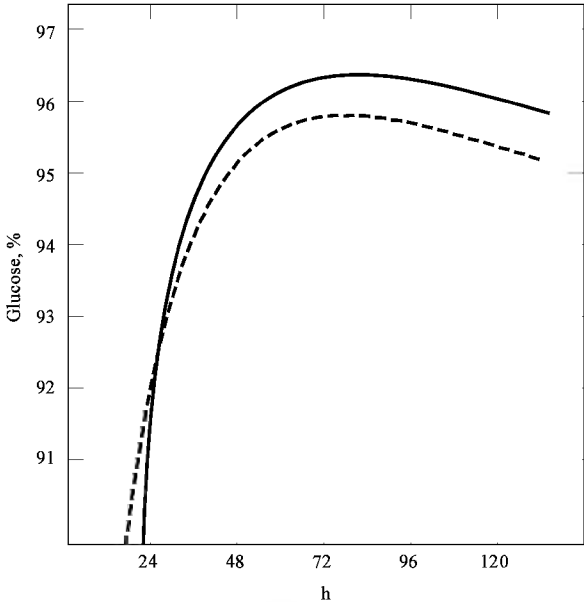


Fig. 10. Saccharification of starch using amyloglucoside (AMG); (—) with pullulanase; (---) without pullulanase. Initial dry substance (DS) of 28° at 60°C, pH 4.3.

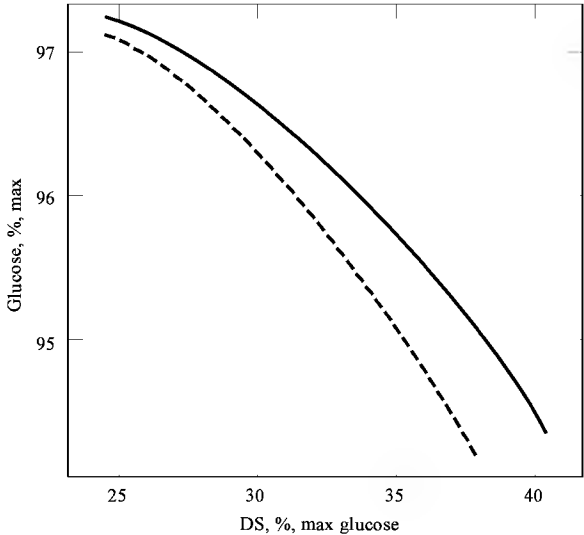


Fig. 11. Effect of dry substance (DS) on maximal obtainable ‰ glucose; (—) AMG and pullulanase; (---) AMG.

The purification of saccharified starch depends on the raw material used, and may be different from plant to plant. When the starch slurry is liquefied in a jet cooker the saccharification process is carried out at 55–65°C, pH 4–4.5, for 24–72 hours. The subsequent steps consist of filtration or centrifugation, ion exchange, isomerization, treatment with activated carbon, and evaporation to form a storage-stable product.

The minor content of impurities found in the starch slurry are connected to the starch granules themselves. To facilitate the purification of the starch during filtration, cellulases, pentosanases, glucanases, proteases, and pectinases are sometimes used. Wheat starch is known to form precipitates or hazes that are difficult to filter. Arabinoxylan, pentosanes, and lysophospholipids are claimed to be responsible for this problem (73).

Glucose Isomerization. Enzymatic isomerization of glucose to fructose provides a real alternative to sugar (sucrose) derived from sugarcane or sugarbeets. The commercial product obtained is known as high fructose corn syrup (HFCS). Two grades of the syrup have become established on the world market, HFCS-42 and HFCS-55, which contain 42 and 55‰ fructose on dry substance basis. These products account for over one-third of the caloric sweetener market in the United States.

Glucose [50-99-7] can be isomerized to fructose [57-48-7] by a reversible reaction. The equilibrium point for glucose⇌fructose is 50‰ under industrial conditions, and the reaction is slightly endothermic (74). The isomerization reaction can only be conducted economically by using immobilized enzymes, and reaction parameters in this system have to be adjusted carefully in order to obtain the optimal yield of fructose; eg, pH approximately 7.5 or higher, and temperatures of between 55 and 60°C to ensure high enzyme activity and stability. Under these conditions glucose and fructose are rather unstable and decompose easily to organic acids and dark-colored by-products. To overcome these problems, the reaction time is kept short by passing glucose continuously through a column of immobilized glucose isomerase.

The Immobilized Enzyme System. The glucose isomerases used are immobilized and granulated to a particle size between 0.3 and 1.0 mm. The enzyme granulates must be rigid enough to withstand compaction when they are packed into the column. Ca²⁺ acts as an inhibitor in the system, and therefore calcium salts need to be removed from the feed syrup. Conversely, Mg²⁺ acts as an activator, and magnesium salts are added to the feed syrup.

When selecting a suitable feed syrup, the main criteria are optimization of enzyme productivity and minimization of the formation of by-products. Typical feed syrup specifications are shown in Table 5. Higher syrup concentration and higher viscosity results in a reduced isomerization rate due to diffusion resistance in the pores of the immobilized enzyme. A deaeration step is desirable to remove dissolved oxygen that would otherwise increase the formation of by-products. The pH is adjusted to the optimum level for the productivity of the enzyme.

Table 5. Feed Syrup Specifications

Specification	Value	
temperature, °C	55–60	
pH	7.5–8.0	
dry substance by weight, ‰	40–50	
glucose content, ‰	>95	
SO ₂ , ppm	0–100	
calcium ion, ppm	< 1	
MgSO ₄ ·7 H ₂ O (activator); g L	0.15–0.75	
conductivity, μS cm	< 100	
uv absorbance, 280 nm	<0.5	

During operation, the immobilized enzyme loses activity. Most commercial enzymes show decay as a function of time (Fig. 12). The glucose isomerase in a reactor is usually replaced after three half-lives, ie, when the activity has dropped to around 12.5‰ of the initial value. The most stable commercial glucose isomerases have half-lives of around 200 days in practical use. To maintain the same fructose content in the finished syrup, the feed-flow rate is adjusted according to the actual activity of the enzyme. With only one isomerization reactor in operation, the result would be excessive variations in the rate of syrup production. To avoid this, several reactors at different stages in the cycle of enzyme decay are operated in combination.

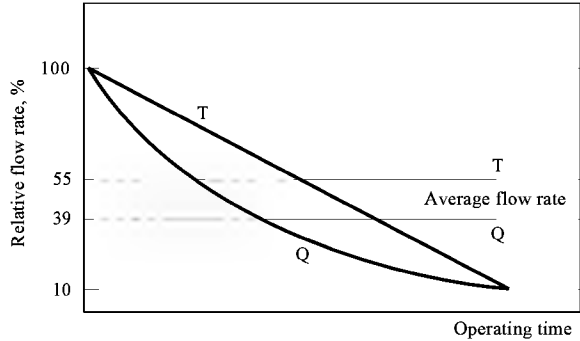


Fig. 12. Activity (syrup flow rate) vs operating time for typical immobilized isomerases, Sweetzyme T and Q.

Reactor design for glucose isomerization in the United States has been documented (75). The diameter of the reactor is normally between 0.6 and 1.5 m. Typical bed height is 2–5 m. The ratio between the bed height and diameter of a reactor should be at least 3:1 to ensure good flow distribution. Plants that produce more than 1000 t of HFCS per day, based on dry matter, use at least 20 separate reactors.

Enzymes in Textile Finishing. To prevent the warp threads, ie, those running along the length of the fabric, from breaking during weaving, the thread is coated with an adhesive substance called a "size." Many different compounds have been used, but since the 1900s starch has been the most common sizing agent. After weaving, the size must be removed to prepare the fabric for bleaching or dyeing, ie, finishing.

Enzymatic desizing is one of the oldest nonfood applications of commercial amylases. Another type of enzyme, microbial cellulases, has developed within the textile industry as a tool for fabric finishing, in particular for denim garment finishing. Cellulases can achieve the fashionable worn look traditionally obtained by the abrasive action of pumice stones, ie, stone-washing.

Cotton Textiles. Starch-splitting enzymes are used for desizing cotton textiles due to their high efficiency and specific way of desizing woven fabric without harmful effects on the yarn. Recommendations on how to desize textiles depend on the type of equipment used, and exact recommendations do not exist. Some general guidance has been published, and a brief guide is given here.

Desizing on a jig is a simple method where the fabric is transferred from one roll to another through a bath. The sized fabric is prewashed in boiling water, whereby the starch is gelatinized. The fabric then goes through an impregnation stage before a thermostable amylase is added. Before adding the enzyme, the desizing liquor is adjusted to pH 6.5–7.0 and a temperature in the range 60–80°C. At the end of the process, the temperature is raised to 90–95°C. The dextrins formed are then removed by washing at 90–95°C for 120 seconds. The jig process is not suitable for continuous high speed

operation, where the enzyme reaction time can be reduced to between 15 and 120 seconds.

Desizing on pad rolls is a continuous process with regard to the passage of the fabric, but a holding time of 2–16 h at 20–60°C is required. A mesophilic α -amylase is used before the size is removed in a wash chamber.

Desizing in a steam chamber is a fully continuous process. The desizing reaction is performed in a steam chamber at 95–100°C. This demands the most temperature-stable amylase available.

Denim Finishing. Most denim garments are treated in laundries before reaching consumers. The garments are given a fashionable appearance such as the stone-wash or acid-wash finishes.

Stone-washing is carried out by lightweight pumice stones that are put into industrial laundry machines with the jeans. The stones rub against the denim and remove some of the dye. However, too much abrasion from stones can damage the fabric, particularly hems and waistbands.

Enzymatic stone-washing is performed either entirely without stones or sometimes by a combination of stones and enzymes. Cellulases are used to attack the surface of the cellulose fiber, but leave the interior intact. Denim garments are dyed with indigo blue, which stays on the surface of the yarn. The cellulase partly hydrolyzes the surface of the fiber, and the indigo blue is partly removed. Either neutral-type cellulases acting at pH 6–8 or acid-type cellulases acting at pH 4–5 are used for these processes.

A typical enzymatic stone-washing process (76) is as follows: load garments into industrial laundry machine, add water, and heat to 50–60°C. Adjust pH to 6.0 with acetic acid or buffer. Desize garment with α -amylase for 10–15 min, and drain water. Add new water, heat to 55–60°C, adjust pH to 6.5–7.0, and add cellulase. Tumble for 20–90 min, drain, rinse twice, and dry.

Backstaining occurs when indigo blue dye that has been lifted off the garment is redeposited onto it. In denim finishing, this effect has to be taken into account carefully. Backstaining depends, among other things, on the pH of the wash liquor. At low (4–6) pH values backstaining is relatively high. However, it is significantly lower in the pH range around neutral. Therefore, neutral cellulases result in a minimum of backstaining and a better stone-wash finish is obtained. The possibility exists to obtain variations in the color and contrast of the denim by using acid cellulases, neutral cellulases, or a combination of both.

Compared with pumice stones, cellulases work without damaging washing machines or garments, there is no need for disposal of the used stones, and the quality of the wastewater is improved. In addition, the labor-intensive job of removing the dust and small stones from the finished garment is eliminated. The overall economics of the process, the higher number of denim garments in each wash load, shorter treatment time, and environmental advantages have made enzymatic stone-washing, or biostoning, a process with great future potential (76).

Biopolishing of Cotton Fabrics. The prevention of pilling, ie, the accumulation of balls of fluff on fabric, and improvement of the smoothness and softness of cotton fabrics are of interest in the textile industry.

In the case of softness, conventional softeners are inclined to be washed out and often make fabrics feel greasy. These problems can be overcome using mixtures of cellulases. The process is referred to as biopolishing. The softness obtained through biopolishing is washproof and nongreasy. Cellulases also have a permanent effect on reducing the tendency to pilling. The enzymatic action only affects the cellulose part of mixed fibers and yarns. The enzyme hydrolyzes the microfibrils protruding from the surface of yarn because these are most accessible. As a result, the microfibrils become weakened and tend to break off. This gives a smoother surface. The improvements in the fabric are obtained with a limited reduction in the bursting strength and no detrimental effect on the fabric's ability to absorb water (77).

Enzymes for Silk Degumming. Raw silk thread and raw silk fabrics must be degummed to remove sericin, a protein substance that covers the fibroin, ie, the silk fiber. Traditionally, degumming is performed using alkaline soap. However, proteolytic enzymes provide an alternative. With enzymes, there is no risk of damaging the fibroin, of excessive degumming, or of uneven dyeing resulting from residues of soap. The cost of treating the wastewater is reduced as well.

Enzymes in the Tannery. The processing of skins and hides for leather (qv) has been based on enzymes ever since 1908 when Otto Rühm patented the first standardized bate based on pancreatic enzymes (78). Leather chemistry research helped to improve understanding of the bating process, and at the same time spurred on developments to improve leather processing (79).

Stages involved in the processing of hides to leather include curing, soaking, liming, unhairing, deliming, bating, pickling, and tanning. The main benefits of using enzymes during the different stages of leather manufacturing are reduced process time, increased opening up of fibrous structure, cleaner surface, increased softness, improved area yield, and reduced need for chemicals.

The animal hide is composed of three main layers. The grain is the outer layer where the hairs are embedded. This layer also contains the small glands producing oils and sweat. The corium represents the principal part of the hide. To convert raw hide into leather, a great deal of interfibrillar material must be removed from this layer. The proteoglycan dermatan sulfate is important because it can be regarded as a kind of cement binding collagen fibrils. The connective tissue on the underside of the hide contains the blood vessels, fat tissue, and some meat proteins. The proteins to take into consideration are collagen, keratin, and some glycoproteins. Acidic polysaccharides attached to a protein core are another important group of substrates.

Cured hides must be properly soaked to obtain satisfactory rehydration and removal of unwanted material. Interfibrillar proteins should be degraded in order to increase water uptake. Bacterial proteases and pancreatic proteases are normally preferred, and are compatible with most tannery chemicals used in soaking, ie, most surfactants and preservatives containing sodium chlorite.

Lipids have been studied more intensively for applications in tanneries. Industrial lipases can be used for degreasing (80).

Enzymes for Liming and Bating. An important discovery (79) for the leather industry was the close relationship between the amount of dermatan sulfate removed and the degree of opening up. Until this discovery, the use of enzymes during the liming and unhairing steps was not considered worthwhile.

During liming, the highly charged dermatan sulfate glycosaminoglycan side chains are split from the proteoglycan back-bone because of the strongly alkaline conditions created by the lime. Thereby a sheath of high ion-density is removed from nearly every collagen fibril. This noncollagenous protein core of the proteoglycan can also be degraded by alkaline-stable proteolytic enzyme preparations. This application is known as enzyme-assisted liming. One example is the use of NUE 0.6 MPX (81) (Novo Nordisk A/S) which provides a number of options, ie, accelerating the unhairing process, improving the opening up of grain leather, reducing the requirement for sulfides by up to 40% (79), reducing the growth and mottle of pelt, increasing the strength of grain leather, and increasing the area yield.

Prior to the bating process, the hides are delimed with ammonium sulfate and/or ammonium chloride. Proteases are then applied. The early preparation proposed by Rühm was pancreatic trypsin. The use of a bating enzyme makes the hides soft and supple to prepare them for tanning. A new microbial protease, Pyrase 250 MP (82) (Novo Nordisk A/S) has been found to be a promising substitute for pancreatic trypsin [9002-07-7], which is more expensive because it must be extracted from pancreatic glands.

Future Developments. Leathermaking is a labor-intensive industry with an eye on costs and improved productivity. Enzymes increase output without requiring investments in new equipment. A large amount of organic waste originates from the beamhouse of a tannery (83), and as environmental awareness grows and regulations become tougher, enzymatic processes may provide ways of treating tannery waste such as chrome shavings and fleshings. Enzymes may also be used to replace chemicals, eg, a degreasing process based on solvents may be substituted by an enzymatic process that can be performed in water.

Enzymes in Pulp and Paper Production. Enzyme-modified starch has been used for adhesives to strengthen paper base and for surface coating. Developments since the late 1980s of further uses of enzymes in papermaking include pitch control and bleach boosting. (see PAPER; PULP).

Pitch Control. Resinous constituents of wood cause problems in paper machines by sticking to the rollers and causing spots or holes in the paper; the worst cases cause paper webs to rupture. Costly stoppages, wastages, and quality problems because of these resinous substances can be avoided by using lipases (84).

Triglycerides are important constituents of resin. In softwood, the triglycerides account for 20–40% of total resin content, and in hardwood, 40–50%. The paper industry uses the term pitch for resins that create problems in paper machines. Traditionally, pitch is controlled or reduced by aging the wood, by use of chemicals to avoid deposits on the rolls, or by intensive washing of the pulp. All these methods add to the cost of paper production. An alternative is to add a lipase to the pulp in a reaction lasting about one hour with the help of agitation. Results from Japanese paper mills show substantial

reductions in pitch-related problems when using a lipase. The lipase is now used regularly to treat the groundwood pulp for the production of newsprint (85).

Bleach Boosting. The lignin residues in chemical pulp have a strong tendency to turn yellow or brown when exposed to light and heat. Residual lignin is traditionally removed in a series of bleaching steps using a combination of chlorination and extraction. The effluent from bleaching plants is of great environmental concern because of the content of chlorinated organic substances. Residual lignin in pulp binds to hemicellulose; enzymes capable of opening up the hemicellulose structure of the pulp have been shown to facilitate the removal of lignin. During the bleaching stage, bleaching chemicals can attack any residual lignin much more effectively if the pulp has been treated with enzymes first. The first experimental industrial enzyme preparations for bleach boosting are xylanases; when used in dosages of approximately 1 kg t of pulp, the desired brightness level of the finished paper can be obtained with one-third less chlorine (86).

Enzymes in the Animal Feed Industry. Many feed components are not fully digested by animals. The direct addition of enzymes to feed can enhance the digestibility of these components. Enzymes have proven to be a successful tool in allowing feed compounders to develop feeds with unconventional and inexpensive formulations. Enzymatic processing of low cost raw materials such as cereals, beans, or seeds allow them to act as substitutes for high quality feed components for young animals. Another area of application is silage production, which can be improved significantly by the combined action of industrial enzymes and lactic acid bacteria.

Enzymes should be added to the feed together with the pre-mix. Granulated enzyme products may readily be mixed with feed components, as they are based on normal feed components such as wheat or soy grits. A wide range of enzyme products are available. Enzyme products should contain specific enzyme activities necessary to degrade specific substances such as glucans, starch, protein, pectin-like polysaccharides, phytic acid, raffinose, stachyose, hemicellulose, and cellulose.

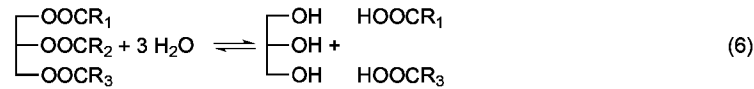
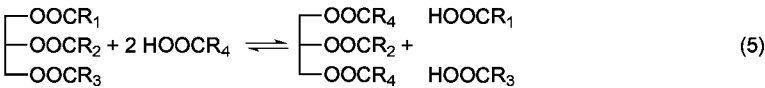
The effect of enzymes, recorded in many feed trials, includes increased final weight of the animal, better feed utilization, improved feed conversion ratio (FCR), more homogeneous production, reduced mortality, and a reduced amount of sticky droppings (in chickens) (87).

The earlier a piglet or calf can be weaned, the better from the breeder's point of view. Great amounts of milk powder are being used in feed milk replacers; with high milk prices, there is an interest in replacing at least part of the milk powder with vegetable protein. Using enzymatic modification, it is possible to make soy protein or rape seed protein perform similarly to milk with regard to nutritional properties and functional properties like solubilization and emulsification. Such modifications are made by using plant cell-wall degrading enzymes, proteases, and specific carbohydrases.

Enzymatic Modification of Lipids. The value of enzymes as biosynthetic agents has been recognized for many years, particularly in the field of lipids. Because of the highly selective mode of action and the ability of specific enzymes to catalyze reactions in organic–aqueous interfaces, enzymes are useful in synthetic organic chemistry. To enhance stability and the rates of reaction, immobilized enzymes are often used.

Concentrated efforts to introduce immobilized lipase technology in the fats and oleochemicals industry have been made since the mid-1980s (88).

A number of specific lipases are used for ester synthesis (eq. 4); transesterification, eg, acidolysis with 1,3 specific lipase (eq. 5); and hydrolysis reactions, eg, with nonspecific lipase (eq. 6).



The components involved in the reactions are oils (triglycerides), glycerol, free fatty acids, esters, and alcohols. Lipases provide an opportunity for the oils and fats industry to produce new types of triglycerides, esters, and fatty acids, and also to make existing products of a higher quality than when using conventional technology. A few examples of products are edible oil that is nutritionally balanced with respect to saturated and unsaturated fatty acids, cocoa butter equivalents, esters for lubricants and cosmetics, monoglycerides as emulsifiers, etc.

Food Applications. A number of features make enzymes ideal catalysts for the food industry. They are all natural, efficient, and specific; work under mild conditions; have a high degree of purity; and are available as standardized preparations. Because enzymatic reactions can be conducted at moderate temperatures and pH values, simple equipment can be used, and only few by-products are formed. Furthermore, enzymatic reactions are easily controlled and can be stopped when the desired degree of conversion is reached.

Dairy Products. Milk is processed into a variety of products. Even in ancient times calf rennet was used for coagulation during cheese production. The milk clotting effect of rennet is due to a specific and limited hydrolysis of the κ-casein surrounding the protein micelles. As a result, the micelles lose their electrostatic charge and are able to aggregate with the help of calcium and phosphate ions to form a network which traps the fat micelles. A gel structure is thus formed. The enzyme present in rennet, chymosin [9001-98-5] (rennin), is extracted from the gastric mucosa of young mammals such as calves and lambs and is a highly specific endoproteinase.

Microbial rennets from a number of producers, eg, Novo Nordisk, Gist Brocades, and Miles, have been available since the 1970s and have proved satisfactory for the production of different kinds of cheese. Their price is considerably lower than that of chymosin. Their properties have proven very similar to those of chymosin (89,90), and only slight modifications of the traditional cheesemaking technique are required in practice.

Microbial rennet may be produced by submerged fermentation of selected strains, eg, the fungus *Rhizomucor miehei*. Various versions of such enzymes have been developed; the principal differences are the thermolability of the enzyme itself. This helps cheesemakers develop their particular type of cheese under local conditions. Products made by recombinant DNA techniques inducing microorganisms to produce chymosin are now being introduced onto the cheese market by Pfizer and Gist Brocades.

In some parts of the world, pepsin is also used to clot milk, but it is much less specific and can give rise to a number of degradation products that tend to taste bitter.

Lactase [9031-11-2] (β-galactosidase) is used to manufacture milk products with a reduced content of lactose, ie, milk sugar, by hydrolyzing it to glucose and galactose. Many people are lactose intolerant and do not have sufficient lactase to digest lactose. By using lactase, lactose can be broken down, and a whole range of lactose-free milk products made. Manufacturers of ice cream, yogurt, and frozen desserts use lactase to improve scoop, creaminess, sweetness, digestibility, and texture of the products.

Two other practical applications of enzyme technology used in dairy industry are the modification of proteins with proteases to reduce possible allergens in cow milk products fed to infants, and the hydrolysis of milk with lipases for the development of lipolytic flavors in specialty cheeses.

Baking. Flour contains enzymes, the most important of which are amylases and proteases. However, the quantities of these enzymes are not always ideal for baking purposes, and supplementary enzymes are often added. Many potential applications of enzymes are being investigated to improve the properties of bread. Traditional applications of enzymes are for improvement of the dough, loaf volume, crumb structure, and shelf-life. The enzyme products used are free-flowing microgranulates that are easy to handle and freely mixed with flour.

To standardize the α-amylase content of flour, a fungal α-amylase is used. Amyloglucosidase [9032-08-0] is used to break down starch, oligosaccharides, and dextrans into glucose to develop crust coloring and, together with fungal α-amylase, for stable chilled or frozen doughs. Fungal α-amylase also improves dough-handling, crumb structure, and loaf volume.

A pentosanase that works on the gluten–pentosan fraction of flour results in easier dough-handling and improved crumb structure. Pentosanases

are able to replace 50–100% of the emulsifiers used, ie, an additive is replaced by a more natural ingredient.

A neutral bacterial endoprotease can be used to weaken the gluten in wheat flour, if necessary, or to provide the plastic properties required in a dough used for biscuits.

A bacterial maltogenic amylase preparation has unique qualities as an antistaling agent. Additionally, it cannot be overdosed, which means that there is no risk of obtaining sticky doughs and a gummy crumb structure.

Brewing. The malting of barley to produce enzymes is one of the central steps in the brewing process (see BEER; BEVERAGE SPIRITS, DISTILLED). If too little enzyme activity is present during malting or mashing the extract yield is low, wort separation takes longer, the fermentation process is slower with less alcohol being produced, beer filtration rate is reduced, and beer flavor and stability is inferior. Industrial enzymes are used to supplement malt enzymes in order to prevent these problems. Industrial enzymes are also used to give a better liquefaction of adjuncts, produce low carbohydrate beer, shorten beer maturation time, and produce beer from cheaper raw materials.

Malt is the traditional source of α -amylase for the liquefaction of adjuncts. The action of α -amylase ensures a simpler liquefaction stage and a reduced process time. There is a strong trend to use heat-stable α -amylase preparations, eg, Termamyl, that are much more stable than malt's own α -amylase. This allows the various malt enzymes to be preserved for the saccharification process where they are used more fully and they are not otherwise inactivated during liquefaction and go to waste; this rationalizes brewhouse operations, gives a better wort, and ultimately better beer. The use of heat-stable α -amylase also eliminates the malt from the adjunct cooker, giving a smaller adjunct mash, and the brewer has greater freedom to balance volumes and temperatures in the mashing program, problems for many brewers who use a high adjunct ratio.

Traditionally, the use of barley has been limited to 10–20% of the grist when using a good-quality malt. When going to higher levels of replacement or when using a lower quality malt, processing becomes more difficult. In these cases, the mash will need to be supplemented with extra enzyme activity in order to be able to maintain performance. The enzymes are added individually according to need or at mashing-in as a malt-equivalent blend of α -amylase, β -glucanase, and protease.

Wort separation and beer filtration are two common bottlenecks in the brewing process. Poor lautering not only causes a loss in production capacity, but can also lead to losses in extract yield. Furthermore, a slow lautering negatively affects the quality of the wort, which may give beer filtration problems and problems with the flavor and stability of the beer.

A thorough breakdown of β -glucans and pentosans during mashing is essential for fast wort separation. Undegraded β -glucans and pentosans carried into the fermenter reduce the beer filter capacity and increase the consumption of kieselguhr. A wide selection of β -glucanase–pentosanase preparations for use during mashing or fermentation maturation are available to solve these problems.

Small adjustments in fermentability can be obtained by adding a debranching enzyme or fungal α -amylase at mashing-in, or by adding a fungal α -amylase at the start of fermentation. Beer types with a very high attenuation can be made with saccharifying enzymes, eg, fungal α -amylase to produce maltose and dextrins for the main part, and amyloglucosidase to produce glucose from both linear and branched dextrins.

If the yeast does not get enough free amino nitrogen, the fermentation will be poor and the beer quality inferior. A neutral bacterial protease added at mashing-in can be used to raise the level of free amino nitrogen. This is useful when working with poorly modified malt or with high adjunct ratios.

Diacetyl is formed by a nonenzymatic oxidative decarboxylation of α -acetolactate, produced by yeast during primary fermentation. The diacetyl is removed again by yeast during the maturation stage by conversion to acetoin, which has a much higher flavor threshold value. By adding α -acetolactate decarboxylase [9025-02-9] at the beginning of the primary fermentation, it is possible to bypass the diacetyl stage and bring about the formation of acetoin during primary fermentation. This makes it possible to shorten or completely eliminate the maturation period (Fig. 13).

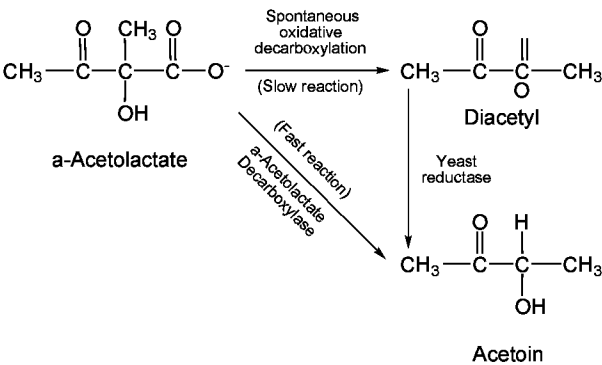


Fig. 13. Removal of α -acetolactate during fermentation.

Protein Modification. The hydrolysis of proteins with enzymes is often an attractive means of obtaining better functional and nutritional properties in food proteins. Enzymes are applied to food proteins for manufacturing new and valuable products from vegetable origin or from animal proteins that are present in by-products, eg, from slaughterhouses. Many different protein raw materials are used with different purposes in mind. Extraction processes with enhanced yields include soya milk, scrap meat recovery, bone cleaning, gelatin, fish meat stick-water, rendering of fat, and deskinning of fish roe. Processes for producing new and promising food ingredients include isoelectric soluble soya protein, egg white substitute from soya protein; emulsifier of soya protein, soluble wheat gluten, foaming wheat gluten, blood cell hydrolyzate, whey protein hydrolyzates, casein hydrolyzates, soluble meat proteins, and gelatin hydrolyzates. The characteristics of some commercial enzyme products used for the industrial conversion of food protein products are shown in Table 6.

Table 6. Commercial Proteolytic Enzymes^a

Product name ^b	Origin	Activity, Anson units g ^c	Practical application	
			pH	°C
Alcalase	<i>Bacillus licheniformis</i>	2.4	6–10	10–80
Esperase	<i>Bacillus lentus</i>	2.4	7–12	10–80
Neutrase	<i>Bacillus amyloliquefacis</i>	0.5	6–8	10–65
Rennilase ^d	<i>Mucor meihei</i>		3–6	10–50
Trypsin ^e	pancreatic	3.3	7–9	10–55

^a Liquid form unless noted.

^b Novo Nordisk trade names.

^c One Anson unit is the amount of enzyme that, under standard conditions, digests hemoglobin at an initial rate, liberating per minute an amount of TCA-soluble product which produces the same color with phenol reagent as one milliequivalent of tyrosine (91).

^d Liquid or granulate form; Anson units standardized in milk clotting activities.

^e Granulate form.

The enzymes used on proteins are proteases that cleave peptide bonds by a hydrolysis reaction. The hydrolysis parameters are a percent protein (S), enzyme–substrate ratio in activity units per kg protein (E/S), pH, and temperature. In addition, the specificity and properties of the enzyme itself are taken into account to determine the course of a reaction on a given protein. The degree of hydrolysis is an important quantitative measure used to assess a proteolytic reaction. This is calculated by determining the number of peptide bonds cleaved and the total number of peptide bonds in the intact protein. The degree of hydrolysis of enzymatically treated proteins can be used to indicate properties of relevance to food applications. It is therefore of the utmost importance that the degree of hydrolysis be measured while the reaction is occurring. Only in this way is it possible to stop the reaction at a definite point, when the desired property of the product has been obtained. The whole subject of methods for monitoring protease reactions has been reviewed (92). Relatively simple analytical tools are often used, eg, pH-stat, pH-drops, osmometry, viscosimetry, and chemical determination of free amino groups.

As a result of the enzymatic degradation of proteins, key indexes change, ie, protein solubility indexes (PSI), peptide chain length (PCL), and protein solubility in 0.8 M TCA (TCA-index) (Fig. 14). Unpleasant bitterness was once a problem for some protein hydrolyzates. This problem can now be overcome by proper selection of the reaction parameters and the enzymes used.

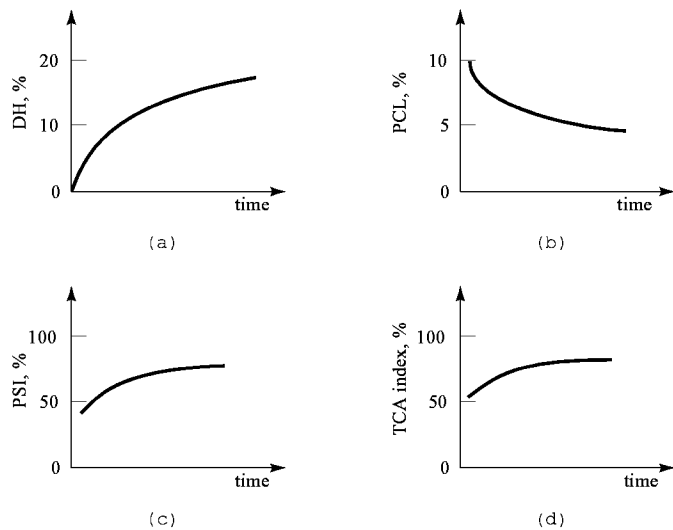


Fig. 14. Time dependence of key indexes during enzymatic hydrolysis: (a) degree of hydrolysis (DH); (b) peptide chain length (PCL) (soluble peptides); (c) protein solubility index (PSI); (d) protein solubility in 0.8 M TCA (TCA index), %.

Bone Cleaning. As an alternative to rendering, an enzymatic process can be used to upgrade fresh bones to valuable products, eg, cleaned bone suitable for gelatin production and a meat protein hydrolysate for the food industry. This process is performed as a two-step enzyme process, ie, scrap meat recovery and bone cleaning.

First, the fresh bone material is crushed and mixed with water to a dry solids content of approximately 25%. Neutrase is added and the slurry is agitated at 60–65°C, pH 6.5–7; inactivation of the enzyme is simultaneously achieved when the reaction is carried out at 60–65°C. To ensure a high yield of hydrolyzate and avoid formation of a bitter flavor, the degree of hydrolysis, according to the TNBS method (92), should be in the range of 5–10%. To defat the hydrolyzate, it is centrifuged at 90°C. The fat is a valuable by-product that is further refined for use in foods. The protein solution may be concentrated or dried, and is used as an ingredient in the meat industry or as a clear soluble meat extract for soups and seasonings.

Bone cleaning is the second stage of enzymatic extraction. The solid bone fraction from the first separation is mixed 1:1 with hot water (65–75°C) and treated with alkaline-type proteases. After a reaction time of one hour, the bones are separated and washed with water. The cleaned bones make an excellent raw material for the production of gelatin.

Modification of Wheat Gluten. Wheat gluten hydrolyzate can be used to make protein-enriched foods and drinks. A completely soluble hydrolyzate is desirable for this application. High protein solubility can be obtained at higher degrees of hydrolysis, as shown in Figure 15. A degree of hydrolysis of about 10% results in solubilization of more than 90% of the gluten. A 100% soluble, blandtasting wheat gluten hydrolyzate with high yield can be recovered by centrifugation and concentration. Inactivation of the protease is carried out during hydrolysis if the reaction is carried out at a temperature above the denaturation temperature of the enzyme (93). Examples of hydrolysis curves are shown in Figure 15. They are plotted by using a pH-stat for the monitoring of the progress of the reaction.

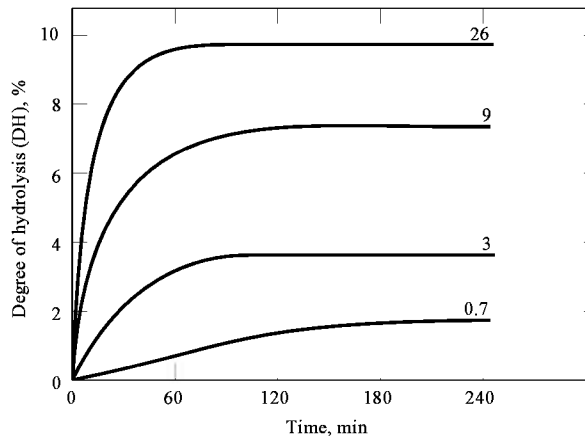


Fig. 15. Enzymatic hydrolysis of wheat gluten at 72.5°C and pH 7.5 by an alkaline protease from *Bacillus licheniformis*. The numbers on the curves are enzyme–substrate ratios (E/S) in activity units (AU) kg of protein where S = 7.4% (N x 5.7).

A protein ingredient with good whipping properties can be used in baked goods and for different types of candy. The optimal whipping properties of wheat gluten are obtained at a degree of hydrolysis of 2–3%. The active whipping protein is recovered by centrifugation and drying.

Extraction Processes of Material. Many ingredients used by the food and brewing industries are produced by extraction from plant matter.

Examples include protein, starch, sugar, fruit juice, oil, flavor, color, coffee, and tea. These are all found in the cells of plant matter, ie, seeds, fruits, etc.

An important development is the degradation of very complex polysaccharides found in the cell walls of unligified plant matter. These cell walls are composed of cellulose fibers to which strands of hemicelluloses are attached. The fibers are embedded in a matrix of pectic substances linked to a structural protein. Enzyme preparations capable of attacking plant cell walls contain different enzyme activities, eg, pectinase, hemicellulase, and cellulase. Conventional enzyme products within these groups are, however, unable to degrade completely the rhamnogalacturonan backbone of the pectic substances. The enzyme complex SPS-ase, named after its intended commercial use as a soya polysaccharide-degrading enzyme, has been produced from a selected strain of *Aspergillus niger* (94). This enzyme preparation contains 10–15 different enzyme activities that offer the possibility of producing different varieties of commercial enzyme products designed for various applications.

Pectinases have been used in fruit juice processing since the 1930s. They are used regularly when making juice from almost all types of fruit and berries, and the use of cell-wall degrading enzymes is also found in winemaking. The enzymes are used to improve the yield of juice, liquefy the entire fruit for maximal utilization of the raw material, improve color and aroma, clarify juice, and break down all insoluble carbohydrates like pectins, hemicellulose, and starch. For the clarification of juice, a mixture of pectinases are required, ie, pectin transeliminase (PTE), polygalacturonase (PG), and pectin methylesterase (PE). Araban is a polysaccharide with a high molecular weight, which may cause haze problems in concentrates. Therefore, apart from the pectinase activities mentioned, clarification enzymes should also contain a substantial amount of arabanase side activity.

Oil from rape seed, coconut, corn germ, sunflower seed, palm kernel, and olives is traditionally produced by a combined process using pressing followed by extraction with organic solvents. Cell-wall degrading enzymes may be used to extract vegetable oil in an aqueous process. They break down the cell-wall structure and release the oil. This concept is already in commercial use in connection with olive oil processing, and has been thoroughly investigated for rape seed oil extraction (95).

Economic Aspects

Worldwide consumption of industrial enzymes amounted to approximately \$720 million in 1990; about one-third was accounted for by the U.S. market. Estimation of worldwide consumption is difficult because official production figures are scarce. A relatively large portion of the production of starch-processing enzymes is for internal consumption. Furthermore, the currency used for the estimation also influences the result considerably.

The growth in volume of the enzyme business from 1980 to 1990 is estimated to be 5–10% per year. The estimated worldwide enzyme sales per industry are shown in Table 7. The detergent and starch conversion industries are by far the most important, and account for 60% of total enzyme sales. Five principal industries account for around 85% of enzyme sales, whereas the remaining sales are spread over many different industries.

Table 7. Estimated Worldwide Enzyme Sales by Industry, 1990

Industry	Sales, 10 \$
detergent	300
starch	125
dairy	80
textile	60
alcohol	45
other	110

Three proteases account for almost all sales to the dairy industry, ie, chymosin extracted from calves' stomachs, chymosin produced by fermentation, and substitutes also produced by fermentation. Four different types of enzyme are used in the detergent industry, ie, proteases, amylases, cellulases, and lipases. Cellulases and lipases have only recently been introduced. Table 8 shows a breakdown of estimated worldwide sales consumption by product types.

Table 8. Estimated Worldwide Enzyme Consumption by Product Type, 1990

Enzyme	Consumption, 10 t
protease ^a	309
rennet (animal and microbial)	74
glucose isomerase	41
glucoamylase	75
amylase (other than glucoamylases)	112
cellulase	55
lipase	11
papain	8
invertase	8
pectinase	7
other	20

^a Includes everything except rennet and papain; bulk share is in the detergent industry.

Environmental and Safety Aspects

The industrial use of microbial enzymes produced by modern biotechnology is an important contribution to the development of green technology. Enzymes have a positive impact on the environment because they replace conventional chemical-based technologies and conventional energy-intensive manufacturing processes, originate from natural biological systems, are totally biodegradable, and leave no harmful residues.

The safety and environmental impact of the production of industrial enzymes can be evaluated on three different levels, ie, the potential risk if the microorganisms, their products, or both are released into the environment; the possible health hazards to staff working with the microorganisms, their products, or both; and safety when products are used by the consumer.

Enzymes are totally biodegradable, and their release into the environment does not cause problems. The release of the production organism itself is controlled by two categories of safety measures which are complementary. The first is physical containment in a fermenter system and recovery plant with a high standard of hygiene. The second is biological containment. Being specially bred, either by traditional techniques or by modern genetic engineering techniques, to produce one specific substance, the production organisms are adapted to grow optimally only under the defined conditions during fermentation. The growth of strains of production organisms in nature is handicapped in comparison with microorganisms already existing in the environment. Their chances of survival in the environment are extremely limited.

Like other proteins, enzymes are potential allergens. In addition, proteases may act as skin and eye irritants. However, during the production and handling of industrial enzymes, the occupational health risks entailed by these properties can be avoided by protective measures, and by the form in which

the enzyme preparations are supplied. In order to reduce dust generation, enzymes are supplied as liquids, encapsulates, or immobilized preparations.

To guarantee that enzymes can be used safely by the consumer, microbial enzymes are obtained from nonpathogenic and nontoxinogenic microorganisms grown on raw materials that do not contain compounds hazardous to health. When a new strain is developed, it is checked for key taxonomic characteristics, and appropriate safety tests are performed. For genetically engineered strains, the new genetic properties are carefully described.

Genetically engineered microorganisms can be used under the same conditions of containment, and the same security rules apply as for equivalent, naturally occurring microorganisms. Provided an enzyme is produced by a harmless host, the contained use of recombinant microorganisms does not warrant any special provisions concerning production conditions, worker protection, environmental assessment, field monitoring, or product approval.

Regulatory Aspects. National authorities have preferred to use or adapt existing legislation and regulations. For the adaptation of the food additive regulations to fit the processing aids applications of enzymes, guidance has been available in the recommendations of the Joint FAO WHO Expert Committee on Food Additives (JECFA), and the Food Chemicals Codex (FCC). The enzyme manufacturers associations, the Association of Microbial Food Enzyme Producers (AMFEP) in Europe and the Enzyme Technical Association (ETA) in the United States, work nationally as well as internationally for a harmonization of regulation. The Codex Committee on Food Additives and Contaminants (CCFAC) plays an important role in this work. In the European Community, common guidelines are being developed for the evaluation of enzymes for food and feed uses. This is a result of the increased attention on new biotechnology rather than on enzymes *per se*. The contained use of genetically modified microorganisms in the manufacture of enzymes has been one of the first cases of a product of recombinant DNA technology reaching the industrial marketplace. A working group on food safety and biotechnology is also considering the use of food enzymes made from genetically modified organisms.

Food Enzymes. The source of a food enzyme determines its primary regulatory status. Traditionally, enzymes from edible parts of plants and animals, eg, papain from papaya and chymosin from calf stomach, have been accepted for food use without further evaluation. Although a few plant and animal enzymes still find industrial uses, the majority of food enzymes are produced by fermentation of microorganisms. Three categories of microorganisms have been defined by JECFA for regulatory purposes: (1) those considered to be food stuffs, such as *Aspergillus oryzae*, the most prominent example of this category; (2) those considered harmless contaminants of food, such as *Aspergillus niger*, *Bacillus subtilis*, *Bacillus licheniformis*, and *Saccharomyces cerevisiae*; and (3) all other microorganisms.

A microbial source for a food enzyme must be nonpathogenic and nontoxicogenic. Manufacturers of microbial food enzymes have always selected their production microorganisms from the safe end of the spectrum of available sources. Consequently, a few species have acquired a record of safe use as sources of a wide variety of food enzymes.

If a food enzyme preparation did not contain any material derived from fermentation other than the enzyme protein itself, the evaluation of its safety would be simple, and no toxicological studies would be required. For enzymes from recombinant microorganisms, the primary regulatory status is determined by the host microorganism, the donor organism, and any vectors involved in the genetic transfer. It is important to note that the new technology makes it possible to establish the function of the various segments of inserted DNA in the recombinant. Whenever the function is known, the origin of that segment of DNA becomes irrelevant.

The majority of food enzymes are used as processing aids, and they have no function in the final food. For that reason, they do not need to be declared on the label, and will not be present in the final food in any significant amount. A few enzymes, however, are used both as processing aids and as food additives. When used as additives, they must be declared on the food label using the appropriate class name, eg, preservative or antioxidant; E-number; and generic name, eg, lysozyme or glucose oxidase. AMFEP has defined Good Manufacturing Practice (GMP) for microbial food enzymes. The most important element is to ensure a pure culture of the production microorganism.

Product specifications for microbial food enzymes have been established by JECFA and FCC. They limit or prescribe the absence of certain ubiquitous contaminants such as arsenic, heavy metals, lead, coliforms, *E. coli*, and *Salmonella*. Furthermore, they prescribe the absence of antibacterial activity and, for fungal enzymes only, mycotoxins.

Enzymes are used as feed digestibility enhancers for chicken and pigs. They must comply with purity specifications comparable to food-grade enzyme specifications. European Community (EC) guidelines for the assessment of additives in animal nutrition are being revised to make them applicable for enzymes. Upon completion of these guidelines, the regulatory status of feed enzymes will be established in EC directive 70 524 EEC.

Technical Enzymes. When an enzyme is used for a technical application, ie, industrial but nonfood and nonfeed, its regulatory status is determined by its properties as a naturally occurring substance. These properties determine the classification and consequent labeling in accordance with existing schemes for chemicals. It should be noted that enzymes are not listed as dangerous chemicals.

Enzyme manufacturers have developed formulations that minimize the release of enzyme dust. In the case of liquid preparations, handling precautions recommend users to avoid the formation of aerosol sprays. In all cases, direct contact with the skin or eyes should be avoided. Enzymes have a good record of occupational health and safety.

BIBLIOGRAPHY

"Enzymes and Enzymology" in *ECT* 1st ed., Vol. 5, pp. 735–762, by A. K. Balls, Agricultural Research Administration, U.S. Department of Agriculture; "Enzymes, Industrial" Suppl. 1, pp. 294–312, by G. Reed, Red Star Yeast and Products Co.; in *ECT* 2nd ed., Vol. 8, pp. 173–230, by G. I. deBecze, Schenley Distillers, Inc.; in *ECT* 3rd ed., Vol. 9, pp. 173–224, by D. Scott, Fermco Biochemicals Inc.

1. A. Payen and J. Persoz, *Ann. Chim.* **53**, 73–108 (1833).
2. T. Schwann, *Arch. Anat. Physiol.*, 90–138 (1836).
3. L. Pasteur, *Ann. Chim.* **3e s rie** **58**, 323–426 (1860).
4. F. W hler and A. Liebig, *Ann. Phys.* **41**, 345–351 (1837).
5. J. J. Berzelius, *Jahres-Berichtung* **15**, 237–245 (1836).
6. D. Keilin, *The History of Cell Respiration and Cytochrome*, University Press, Cambridge, 1966.
7. O. Hoffmann-Ostenhof, *TIBS*, 186–189 (Aug. 1978).
8. W. K hne, *Untersuchungen aus dem Physiologischen Institut Heidelberg I*, 291–324 (1878).
9. *Enzymes at Work*, Novo Nordisk A/S, Bagsvaerd, Denmark, 1989, pp. 40–45.
10. C. Dambmann and co-workers, *Dev. Ind. Microbiol.* **12**, 11–23 (1971).
11. **Ger. Pat. 283,923** (1913), Dr. Otto R hm.
12. *Enzyme Nomenclature 1978*, Academic Press, Inc., London, 1979, pp. 3–26.
13. O. P. Ward, ed., *Fermentation Biotechnology: Principles, Processes, Products*, Vol. 1, Open University Press, Milton Keynes, 1989.
14. J. S. Rokem in A. Mizrahi, ed., *Upstream Process: Equipment and Techniques*, Alan R. Liss, Inc., New York, 1988, pp. 49–78.
15. M. H. Siegel, M. Halla le, and J. C. Merchuk in Ref. 14, pp. 79–124.
16. J. J. S. Snell in B. E. Kirsop and J. J. S. Snell, eds., *Maintenance of Microorganisms*, Academic Press, Inc., London, 1984, p. 11.
17. R. H. Rudge in Ref. 16, p. 23.
18. T. G. Watson, I. Nelligan, and L. Lessing, *Biotechnol. Lett.* **6**, 667–672 (1984).
19. J. A. Asenjo in D. L. Pyle, ed., *Separations for Biotechnology*, Vol. 2, Elsevier Applied Science, London, 1990, pp. 11–20.
20. K. H. Brunner and H. Hemfort in Ref. 14, pp. 1–50.
21. J. Shiloach, N. Martin, and H. Moes in Ref. 14, pp. 97–125.
22. M. D. White and D. Marcus in Ref. 14, pp. 51–96.
23. H. Walter, D. E. Brooks, and D. Fisher, eds., *Partitioning in Aqueous Two-Phase Systems: Theory, Methods, Uses and Applications to Biotechnology*, Academic Press, Inc., Orlando, 1985.

24. K. Visuri and co-workers, *Bio/Technology* **8**, 547–559 (1990).
25. E. L. V. Harris and S. Angal, *Protein Purification Methods—A Practical Approach*, IRL Press, Oxford, 1989.
26. P.-E. Albertsson, G. Johansson, and F. Tjerneld in J. A. Asenjo, ed., *Separation in Process Biotechnology*, Marcel Dekker, Inc., New York, 1990, pp. 287–327.
27. R. D. Schmid in T. K. Ghose, A. Fiechter, and N. Blakebrough, eds., *Advances in Biochemical Engineering*, Vol. 12, Springer-Verlag, Berlin, 1979, pp. 41–118.
28. M. F. Chaplin and C. Bucke, *Enzyme Technology*, Cambridge University Press, Cambridge, 1990.
29. **U.S. Pat. 4,318,818** (Mar. 9, 1982), J. C. Letton and M. J. Yunker (to Procter & Gamble Co.); **U.S. Pat. 4,261,868** (Apr. 14, 1981), J. Hora and G. A. A. Kivits (to Lever Brothers Co.).
30. E. Lueck, *Antimicrobial Food Additives*, Springer-Verlag, Berlin, 1980.
31. K. H. Wallhauser, *Praxis der Stabilisation, Desinfektion-Konservierung*, Georg Thieme Verlag, Stuttgart, 1988.
32. WO 90 00593, (to Novo Nordisk A S and Albright & Wilson Ltd.).
33. K. Mosbach ed., *Methods in Enzymology*, Vol. 44, Academic Press, Inc., New York, 1976.
34. *Ibid.*, Vol. 135.
35. *Ibid.*, Vol. 136.
36. *Ibid.*, Vol. 137.
37. S. Pedersen in T. Kobayashi, A. Tanaka, and T. Tosa, eds., *Industrial Application of Immobilized Biocatalysts*, Marcel Dekker, Inc., New York, 1991.
38. L. G. Karlsen, *Use of Semacylase in 6-APA Production from Penicillin V*, publication B 528, Novo Nordisk, Bagsvaerd, Denmark, 1990.
39. *Revised Operating Guidelines 5th Report*, The Soap and Detergent Industry Association, Hayes, Middlesex, UK, 1991.
40. J. Falbe, ed., *Surfactants in Consumer Products*, Springer-Verlag, New York, 1987.
41. J. C. Harris, *Detergency Evaluation and Testing*, Interscience Publishers, New York, 1954.
42. "Durazym: The First Enzyme Designed by Novo Nordisk," *Soap, Cosmet. Chem. Spec.*, (Aug. 1991).
43. H. Andree and co-workers, *J. Appl. Biochem.* **2**, 218–229 (1980); T. Tatara and co-workers, *J. Am. Oil Chem. Soc.* **62**, 1053–1058 (1985).
44. T. Fujii and co-workers, *J. Am. Oil Chem. Soc.* **63**, 796–799 (1986).
45. K. Umehara, and co-workers, *Yukagaku* **39**, 321–326 (1990).
46. **Eur. Pat. Appl. 305,216**, E. Boel and B. Høge-Jensen (to Novo).
47. **Eur. Pat. Appl. 334,462** (Sept. 27, 1989), P. M. Andreoli and co-workers, (to Gist-Brocades NV).
48. **Eur. Pat. Appl. 407,225** (Jan. 9, 1991), A. M. Batenburg and co-workers, (to Unilever PLC).
49. E. Boel and co-workers, *Lipases: Structure, Mechanism and Genetic Engineering*, Vol. 16, GBF Monographs, 1991, pp. 207–219.
50. D. Aaslyng, E. Gormsen, and H. Malmos, *J. Chem. Technol. Biotechnol.* **50**, 321–330 (1991).
51. H. Malmos and co-workers, *Chem. Ind.* **6**, 183–186 (Mar. 19, 1990).
52. **Eur. Pat. EP 0177 165 B1** (Apr. 9, 1986) J. R. Martin and J. R. Nooi (to Unilever PLC).
53. M. Murata and co-workers, *JAOCS* **68**(7) (July 1989).
54. A. Suzuki, *Research Report*, Kao Cooperation Research Laboratories, Beseibutsu, 1989, pp. 100, 104, 105.
55. **Eur. Pat. 0,171,006-8 A** (Feb. 12, 1986), E. Sung and co-workers (to HENKEL KGAA).
56. **Eur. Pat. 0,318,204 A** (May 31, 1989), J. Vandijk, H. Frankena, and K. S. Kielman (to Unilever PLC).
57. **Eur. Pat. 0,414,197 A** (Feb. 27, 1991), L. Hertling, W. Ussat, and G. Waeschenbach (to Benckiser GmbH).
58. Application sheet B 077d-GB 3000, Novo Nordisk A S, May 1990.
59. *Methods for Measuring the Performance of Electric Dishwashers*, publication 436, 2nd ed., International Electrotechnical Commission Standard, 1981.
60. *Deposition on Glassware During Mechanical Dishwashing*, CSMA Designation DCC-05A, Dec. 1981.
61. E. Kissa in W. G. Culter and E. Kissa, eds., *Detergency Theory & Technology, Surfactant Science Series*, Vol. 20, Marcel Dekker, Inc., New York, 1987, pp. 304–307.
62. **Eur. Pat. 271,155 A** (June 15, 1988), W. R. Vandijk (to Unilever NV).
63. L. H. Dalgerd and co-workers, *Inform* **2**(6), 532–536 (1991).
64. **Jpn. Pat. 0 207 7499 A** (Mar. 16, 1990), (to Lion Corp.).
65. **Eur. Pat. 346,136 A** (Dec. 13, 1989), W. R. Vandijk (to Unilever PLC).
66. **Eur. Pat. 346,137 A** (Dec. 13, 1989), J. Klugkist and W. R. Vandijk (to Unilever PLC).
67. C.-Y. Lii and Y.-H. Chang, *Food Reviews Int.* **7**(2), 185–203 (1991).
68. *Use of Termamyl for Starch Liquefaction*, application sheet, Enzyme Process Div. of Novo Nordisk A S, Bagsvaerd, Denmark, 1990.
69. H. Chouvel, P. B. Chay, and J.-C. Cheftel, *Lebensm. Wiss. Technol.* **16**, 346–353 (1983).
70. A. Moellgaard and co-workers, "Continuous Low-Temperature Cooking of Wheat for Production of Ethanol," paper presented at the 7th International Symposium on Alcohol Fuels, Paris, Oct. 20–23, 1986.
71. B. E. Norman, *A Novel Debranching Enzyme for Application in the Glucose Syrup Industry*, Starch Starke **34**, no. 10, 1982, pp. 340–346.
72. J. R. Reichelt in T. Godfrey and J. Reichelt, eds., *Industrial Enzymology*, Macmillan Publishers Ltd., UK, 1983.
73. G. Konieczny-Janda and G. Richter, *Progress in the Enzymatic Saccharification of Wheat Starch*, Starch Starke **43**, no. 8, 1991, pp. 308–315.
74. Y. B. Tewari and R. N. Goldberg, *J. Solution Chem.* **13**, 523–547 (1984).
75. P. H. Blanchard and E. O. Geiger, *Sugar Technol. Rev.* **11**, 1–94 (1984).
76. D. Kochavi, T. Videbaek, and D. Cedroni, *Am. Dyest. Rep.*, 24–28 (Sept. 1990).
77. L. O. Asferg and T. Videbaek, *Softening and Polishing of Cotton Fabrics by Cellulase Treatment*, ITB Dyeing Printing Finishing, 2 90 (1990).
78. **U.S. Pat. 886,411** (1908), O. Ruhm.
79. K. T. W. Alexander, *JALCA*, **83**, 287–316 (1988).
80. J. Christner, "The Use of Lipases in the Beam House Processes," paper presented at the 87th American Leather Chemists Association's Annual Meeting, Waterville Valley, N.H., June 1991.
81. *Application of NUE 0.6 MPX for Liming and Unhairing of Hides*, application sheet, Enzyme Process Div. of Novo Nordisk A S, Bagsvaerd, Denmark, 1991.
82. *Pyrase 250 MP*, product sheet, Enzyme Process Div. of Novo Nordisk A S, Bagsvaerd, Denmark, 1989.
83. *BioTimes* **5**(1), 2–5 (1990).
84. Y. Irie and co-workers, *Enzymatic Pitch Control in Papermaking Systems*, Papermakers Conference TAPPI proceedings, 1990, p. 10.
85. *BioTimes* **6**(1), 8–9, (1991).
86. L. S. Pederson, *On Use of Pulpzyme HLA for Bleach Boosting*, Novo Nordisk A S, Bagsvaerd, Denmark, Sept. 1989.
87. *Biofeed, Novo Nordisk Enzymes for the Feed Industry*, publication B-507 from Enzyme Process Div. of Novo Nordisk A S, Bagsvaerd, Denmark.
88. P. Eigtved in F. B. Padley, ed., *Advances in Applied Lipid Research*, Vol. 1, JAI Press Ltd., London, 1992, pp. 1–64.
89. I. A. Frederick and S. C. Fuller, *Aust. J. Dairy Technol.*, 12–15, (May 1988).
90. G. van den Berg and co-workers, *The use of Rennilase XL for the Manufacture of Gouda Cheese*, Nizo report R 126, 1987.
91. M. L. Anson, *J. Gen. Physiol.* **22**, 79–89 (1939).
92. J. Adler-Nissen, *Enzymic Hydrolysis of Food Proteins*, Applied Science Publishers, London, 1986.
93. H. S. Olsen, *Food Technology International Europe 1988*, Sterling Publications Limited, 1988, pp. 245–250.

94. U.S. Pat. **4,478,939** (Oct. 23, 1984), J. Adler-Nissen and co-workers (to Novo Nordisk A S).
95. H. S. Olsen and F. M. Christensen, Proceedings from the *7th World Congress of Food Science & Technology*, Singapore, Sept. 28th–Oct. 2nd, 1987, pp. 139–146.

General References

J. E. Bailey and D. F. Ollis, *Biochemical Engineering Fundamentals*, 2nd ed., McGraw-Hill Book Co., Inc., New York, 1986, Chaps. 3–4.
A. Fersht, *Enzyme Structure and Mechanism*, 2nd ed., W. H. Freeman and Co., New York, 1988.
T. Palmer, *Understanding Enzymes*, 1st ed., Ellis Horwood, New York, 1991.
D. V. Roberts, *Enzyme Kinetics*, 1st ed., Cambridge University Press, London, 1977.
L. Stryer, *Biochemistry*, 3rd ed., W. H. Freeman and Co., New York, 1988, Chaps. 8–10.
D. Freifelder, *Physical Biochemistry*, 2nd ed., W. H. Freeman and Co., New York, 1982.
R. J. Leatherbarrow, *TIBS* **15**, 455–458 (1990).

Peder Holk Nielsen
Henrik Malmos
Ture Damhus
Boerge Diderichsen
Henrik Kim Nielsen
Merete Simonsen
Hans Erik Schiff
Anders Oestergaard
Hans Sejr Olsen
Peter Eigtvad
Tage Kjaer Nielsen
Novo Nordisk A/S

THERAPEUTIC

Advances in the knowledge of therapeutic uses of enzymes have heightened interest in the manufacture and processing of these macromolecules. Therapeutic uses of proteolytic enzymes were attempted in the early nineteenth century; in 1902, Emmerich first demonstrated the therapeutic application of a nuclease capable of degrading nucleic acids (1). This milestone study opened the way for the use of enzymes, first as crude topical and oral preparations with proteolytic and hydrolytic activity, and later as highly purified proteins for use in cancer chemotherapy, genetic metabolic deficiencies, clotting disorders, and as antidotes to treat poisons, drug toxicities, and kidney failure (2–7). Immobilized, modified, and entrapped enzyme preparations have become available to circumvent some of the side effects inherent in the use of foreign, or immunogenic, proteins.

Physical and Chemical Properties

Enzymes are protein catalysts of remarkable efficiency and specificity. Lipid, carbohydrate, nucleotide, or metal-containing prosthetic groups may be attached to these enzymes and serve as essential components of their catalyses by enhancing specificity and or stability (8–13). Each enzyme has a specific temperature and pH range where it functions to its optimal capacity; the optima for these proteins usually lie between 37–47°C, and pH optima range from acidic, ie, 1.0 in the case of gastric pepsin, to alkaline, ie, 10.5 in the case of alkaline phosphatase. However, enzymes from extremely thermotolerant bacteria have become available; these can function at or near the boiling point of water, and therapeutic use of these ultrastable proteins can be anticipated.

Crude enzyme extracts are often unsuitable for therapeutic uses because of their antigenicity, contamination with endotoxins, and rapid inactivation under physiological conditions or in fluids intended for intravenous infusion over several hours. When the enzyme used is a foreign protein, it can elicit an immune response that alters the clearance rate or induces severe allergic reactions in the host. After an intravenous injection of an enzyme, its activity in plasma decreases with time due to distribution to other fluids and tissues, and as a consequence of proteolysis or excretion. Distribution is related to molecular size, charge, and lipophilicity; surface charges attributable to the availability of free amino, amido, or carboxyl groups may affect the rate of inactivation of some enzymes.

Hydrolases represent a significant class of therapeutic enzymes [Enzyme Commission (EC) 3.1–3.11] (14) (Table 1). Another group of enzymes with pharmacological uses has built-in cofactors, eg, in the form of pyridoxal phosphate, flavin nucleotides, or zinc (15). The synthases, and other multisubstrate enzymes that require high energy phosphates, are seldom available for use as drugs because the required co-substrates are either absent from the extracellular space or are present in prohibitively low concentrations.

Table 1. Therapeutic Enzymes

Enzyme	CAS Registry Number	Enzyme commission number	Catalysis	Use
neuraminidase	[9001-67-6]	3.2.1.18	hydrolysis of terminal acylneuraminyl residues	antineoplastic
ribonuclease	[9001-99-4]	2.7.7.16	RNA → oligoribonucleotides	antineoplastic
L-α-arabinofuranosidase	[9067-74-7]	3.2.1.55	L-α-arabinofuranoside → alcohol + L-arabinose	antineoplastic
brinase	[9074-07-1]	3.4.21.15	fibrinogen → fibrin	fibrinolytic
α-glucosidase	[9001-42-7]	3.2.1.20	D-1, 4α-glucoside → α-D-glucose	metachromatic leukodystrophy
β-glucosidase	[9001-22-3]	3.2.1.21	D-1, 4β-glucoside → β-D-glucose	type A glycogenosis
arylsulfatase	[9016-17-5]	3.1.6.1	phenolsulfate → phenol + sulfate	metachromatic leukodystrophy
α-galactosidase	[9025-35-8]	3.2.1.22	α-D-galactoside → α-D-galactose	Fabry's disease
β-galactosidase	[9031-11-2]	3.2.1.23	β-D-galactoside → β-D-galactose	Fabry's disease
bromelain	[37189-34-7]	3.4.22.4	protein → amino acids and peptides	antiinflammatory
collagenase	[9001-12-1]	3.4.24.3	collagen → amino acids and peptides	dermal ulcers
papain	[9001-73-4]	3.4.22.2	protein → amino acids and peptides	reduction of edema after dental surgery
L-asparaginase	[9015-68-3]	3.5.1.1	L-asparagine → L-aspartic acid + NH ₃	antineoplastic
streptokinase	[9025-51-8]	3.4.22.10	plasminogen → plasmin	thrombolytic
arvin	[9046-56-4]		fibrinogen → fibrin	fibrinolytic
urokinase or tissue plasminogen activator	[9039-53-6]	3.4.21.31	plasminogen → plasmin	thrombolytic
coagulation factor VIII	[9001-27-8]		prothrombin → thrombin	hemophilia
glucocerebrosidase	[37228-64-1]		glycolipid glucocerebroside → α-D-glucose + ceramide	Gaucher's disease
lipase	[9001-62-1]	3.1.1.3	carboxylic ester → alcohol + carboxylic acid	pancreatic deficiency
L-glutaminase	[9001-47-2]	3.5.1.2	L-glutamine → L-glutamic acid + NH ₃	antineoplastic
L-arginase	[9000-96-8]	3.5.3.1	L-arginine → L-ornithine + urea	antineoplastic
L-tyrosinase	[9002-10-2]		L-tyrosine + O ₂ → dihydrophenylalanine + H ₂ O	antineoplastic
L-serine dehydratase	[9014-27-1]	4.2.1.13	L-serine → pyruvate + NH ₃	antineoplastic
L-threonine deaminase	[9024-34-4]	4.2.1.16	L-threonine → 2-ketobutyric acid + NH ₃	antineoplastic
L-tryptophanase	[9024-00-4]	4.1.99.1	L-tryptophan → indole + pyruvate + NH ₃	antineoplastic
deoxyribonuclease	[9003-98-9]	3.1.4.5	DNA → oligodeoxyribonucleotides	chronic bronchitis
trypsin	[9002-07-7]	3.4.21.4	protein → peptides	athletic injuries
chymotrypsin	[9004-07-3]	3.4.21.1	protein → peptides	athletic injuries

superoxide dismutase	[9054-89-1]	1.15.1.1	$\text{O}_2^- + \text{O}_2^- + 2\text{H}^+ \longrightarrow \text{O}_2 + \text{H}_2\text{O}_2$	antiinflammatory
----------------------	-------------	----------	---	------------------

It is essential to maintain high maximal velocities of enzymatic activity for the attainment of optimal therapeutic efficacy. As a general rule, only enzymes whose Michaelis-Menten constants lie between 1–100 μM are effective as drugs (16) because most substrates for therapeutically useful enzymes are present in body fluids and cells at submillimolar concentrations.

Commercial enzymes are available in oral form, sometimes formulated with appropriate stabilizers and excipients. However, such preparations are seldom suitable for parenteral use. Therefore, dry preparations devoid of high salts, excipients, or reducing agents have been adopted for the final formulation of therapeutic enzymes intended for systemic administration. The most commonly used method for the formulation of therapeutic enzymes involves lyophilization in the presence of mannitol and a physiological buffer. Since many enzymes can be denatured by heat even in the dry state, refrigeration or freezing during transit or storage is customary.

The therapeutic utility of an enzyme preparation is largely dependent on its stability as finally formulated. A number of chemical modifications have been employed and include binding to inert surfaces and encapsulation to increase the resistance of these intrinsically labile macromolecules to decomposition. Unfortunately, some of these modifications can interfere with the optimal kinetic performance of the enzyme. The development and deployment of recombinant DNA technologies have made significant contributions toward meeting the goal of mass production of human enzymes for human use. However, it must be stressed that post-translational modifications of a given enzyme may not occur in the foreign organism or cell used to manufacture it. Some of these modifications may prove to be essential for catalytic activity, or for imparting other necessary therapeutic characteristics, and it may be necessary to carry out these modifications by chemical means (17).

Uses

Therapeutic enzymes have a broad variety of specific uses, ie, as oncolytics, thrombolytics, or replacements for inherited deficiencies. Additionally, there is a growing group of miscellaneous enzymes of diverse function.

Oncolytic Enzymes. An early report of cancer chemotherapy using an enzyme, pepsin [9001-75-6], was published in 1922 (18); its clinical use was surrounded by controversy.

Enzymes Degrading Amino Acids. One principal stimulus for the successful enzymic therapy of cancer came from the observation that L-asparaginase from guinea pig serum had striking antilymphoma activity (4). This enzyme decomposes L-asparagine [56-86-0] to L-aspartic acid [73-22-3] and ammonia [7664-41-7]. Subsequent clinical trials with L-asparaginase from bacterial sources, principally *Escherichia coli* and *Erwinia carotovora*, showed pronounced antileukemic activity in the therapy of acute lymphoblastic leukemia (19). The attachment of poly(ethylene glycol) to proteins protracts their plasma half-lives, and reduces their immunogenicity (20); the covalent attachment of poly(ethylene glycol) to L-asparaginase yields a polymer–enzyme conjugate whose plasma half-life ranges from 2 to 3 weeks, with minimal immunogenicity. In contrast, unconjugated bacterial L-asparaginase preparations have plasma half-lives of approximately 24 hours, require repetitive dosing (21,22), are strongly immunogenic, and its use is complicated by the rapid emergence of L-asparaginase-resistant tumor cells. The poly(ethylene glycol) conjugate of L-asparaginase exhibits activity even in patients with non-Hodgkin's lymphoma, a disease ordinarily refractory to the unmodified enzyme (21).

Therapy with L-asparaginase is most successful against tumors exhibiting a deficiency in the synthesis of L-asparagine. Most normal cells exhibit a healthy capacity to synthesize this nonessential amino acid and are not damaged by exposure to L-asparaginase (23). This finding demonstrates that biochemical differences between normal and cancer cells can be exploited for successful cancer chemotherapy.

L-Asparaginase is used for the treatment of appropriate lymphoproliferative disorders; in two clinical trials, L-asparaginase was used in combination with chemotherapy for the treatment of refractory acute nonlymphocytic leukemia in children (24) and adults (25). A moderate efficacy, attributable to the enzyme, was demonstrated in both trials.

A review of the synthetic pathways and requirements for amino acids of eukaryotic cells shows that tumor cells require abundant supplies of L-glutamine, L-arginine, L-cysteine, L-tryptophan, and other amino acids for growth and survival (26). An L-glutaminase–asparaginase preparation from *Acinetobacter*, which hydrolyzes L-glutamine [70-47-3] or L-asparagine to L-glutamic acid [56-84-8] or L-aspartic acid, respectively, shows activity against cultured tumor cells (27–29). In addition, the chemical modification of this protein with succinic anhydride prolongs its plasma half-life in humans (30) in much the same way as the poly(ethylene glycol) modification. L-Arginine desimidase from *Streptococcus faecalis*, mammalian L-arginase coupled to poly(ethylene glycol), and L-arginine decarboxylase, which converts L-arginine to agmatine, have all shown antitumor activity vs experimental neoplasms (27). In a study where 17 commercially available amino acid degradative enzymes were tested, L-asparaginase and L-lysine decarboxylase [9024-76-4], which decarboxylates L-lysine [63-68-3] to form cadaverine [462-94-2], were found to be effective inhibitors of lymphocyte growth, followed by arginase which hydrolyzes L-arginine [7004-12-8] to L-ornithine [70-26-8] and urea [57-13-6], and L-tyrosinase which converts L-tyrosine [63-91-2] to dopaquinone [4430-97-1]; other enzyme preparations were ineffective (31).

Although essential amino acids are required by both host and tumor, deprivation of select essential amino acids for 2–3 weeks is tolerated by the host yet exerts a pronounced antiproliferative effect on the tumor. Thus, treatment of mice with indole-3-alkane-α-hydroxylase [63363-76-8] from *Pseudomonas*, which transforms L-tryptophan [73-22-3] to 3-indolylglyceraldehyde, lowers the concentration of L-tryptophan in plasma, brain, and lungs, and inhibits the growth of a variety of tumors (32–34).

Certain tumor cells cannot synthesize L-methionine from L-homocysteine. For this reason, L-methionine degrading enzymes from *Clostridium sporogenes* and other sources have been used successfully against a variety of murine tumors (32–34). L-Phenylalanine ammonia lyase [9024-28-6], an enzyme which deaminates both L-phenylalanine [63-91-2] and L-tyrosine, is effective in the therapy of transplantable tumors of mice (35,36) but has yet to be studied in a systematic way against human cancer. Antimetastatic activity was observed after treatment of mice bearing Lewis lung carcinoma with L-lysine-α-oxidase (37). Other amino acid-degrading enzymes with oncolytic activity in animal tumors include L-tyrosinase (38); L-serine dehydratase, which decomposes L-serine [56-45-1] to form pyruvate (39); L-threonine deaminase, which deaminates L-threonine [56-45-1] to form 2-oxobutanoate [600-18-0] (40); L-tryptophanase (41); and L-cysteine–cystine degrading enzymes (42).

A different kind of enzyme, translocase [80700-39-6], which transfers a fragment of NAD to the protein–synthesis factor (elongation factor 2), is catalyzed by diphtheria toxin, thereby inhibiting protein synthesis (43). In tumor cells, the rate of protein synthesis is 100 to 1000 times more sensitive to diphtheria toxin than the analogous process in normal cells (41); therefore, diphtheria toxin is selectively toxic to tumor cells.

Enzymes Degrading Macromolecules. Enzymes that degrade macromolecules such as membrane polysaccharides, structural and functional proteins, or nucleic acids, have all shown oncolytic activity. Treatment strategies include the treatment of inoperable tumors with pepsin (1); antitumor activity of carboxypeptidase G₁ (44); cytotoxicity of ribonuclease (45–47); oncolytic activity of neuraminidase (48–52); therapy with neuraminidase of patients with acute myeloid leukemia (53); antitumor activity of proteases (54); and hyaluronidase treatment in the management of human solid tumors (55).

Carboxypeptidases, especially carboxypeptidase G₁ [9054-73-3] from *Pseudomonas*, hydrolyze peptide bonds at the carboxyl terminus of proteins and, in an analogous process, hydrolyze the terminal L-glutamic acid moieties of folic acid, which is naturally polyglutamylated in cells. This action induces a state of folic acid deficiency deleterious to the tumor cell. Systemic administration of carboxypeptidase G₁ also hydrolyzes polyglutamylated methotrexate [82334-40-5], an analogue of folic acid with antitumor activity, in blood but not in cerebrospinal fluid because its molecular mass prevents it from penetrating the blood brain barrier (44). This selectivity prolongs the action of methotrexate in brain while the drug is polyglutamated in other compartments. A pilot human trial of carboxypeptidase G₁ in a patient with a brain tumor showed that the enzyme was excluded from the cerebrospinal fluid by the blood brain barrier. Such selectivity permitted the antifolate, methotrexate [59-05-2] to exert its antiproliferative action on the brain tumor; at the same time the patient's normal tissues were protected from the side effects of the drug (56).

The ribonucleases, a family of enzymes that degrade RNA, are abundant in plasma, urine, and tissues (45). Studies with seminal ribonuclease have

shown that it exerts antitumor activity against several murine plasmacytomas (46); antitumor activity was also observed after administration of a ribonuclease from *Bacillus intermedius* to mice bearing a transplantable lymphocytic leukemia (47).

Neuraminidase removes sialic acid residues from the surface of cancer cells, thereby altering their immunogenicity and in certain cases rendering them susceptible to the host's immune response (48–52). Although this treatment alters the physical, biological, and immunological properties of tumor cells, it does not change the cells' viability. Injection of these treated cells regularly induces a specific immune reaction and improves the survival of animals subsequently inoculated with untreated tumor cells. Treatment with neuraminidase of mice bearing solid metastasizing tumors, ie, Lewis lung carcinoma and mammary carcinoma, yields noteworthy increases in life-span (55,57,58). Significantly longer remissions and survival in leukemic patients were demonstrated when they were immunized with allogenic myeloblasts previously treated with neuraminidase from *Vibrio cholerae* (53).

A second family of carbohydrate-degrading enzymes, the lysozymes, produces synergistic antimetastatic activity when co-administered with cisplatin [15663-27-1] to mice whose primary tumor had been surgically removed (51).

Mucopolysaccharide levels are increased in many cancer cells; this increase is accompanied by a decreased vascular supply. Hyaluronidase [9001-54-1], obtained from bovine testis, dissolves the mucopolysaccharides surrounding a tumor, thereby allowing cytotoxic agents to penetrate the neoplasm with enhanced facility. In clinical studies hyaluronidase was given to patients with malignant tumors, alone or in combination with 5-fluorouracil [51-21-8]; significant decreases in tumor mass were observed (55).

Protease has also been demonstrated to exhibit antitumor activity. Intratumoral microinjection of proteases from *Serratia marcescens* into mice with solid tumors resulted in necrosis and solubilization of the tumor mass (53).

Ricin [9009-86-3], a phytotoxin found in the seeds of the castor oil plant *Ricinus communis*, conjugated to murine monoclonal antibody (Immunogen Corp.), has been approved by the U.S. Food and Drug Administration (FDA) for the treatment of patients with B-cell leukemia and lymphoma (59).

Thrombolytic Enzymes. Although atherosclerosis and the accompanying vascular wall defects are ultimately responsible for such diseases as acute pulmonary embolism, arterial occlusion, and myocardial infarction, the lack of blood flow caused by a fibrin clot directly results in tissue injury and in the clinical symptoms of these devastating diseases (54). Thrombolytic enzyme therapy removes the fibrin clot by dissolution, and has shown promise in the treatment of a number of thrombo-occlusive diseases (60).

The most important use of thrombolytic therapy is in acute myocardial infarction, a leading cause of death in the United States. The fibrin cross-links that form the structural skeleton of a thrombus can be effectively degraded by an endogenous serine protease, plasmin [9001-90-5], which is released following activation of its zymogen precursor, plasminogen [9001-91-6]. Human plasminogen is an inactive plasma protein secreted by the liver and circulating at levels of about 20 mg dL. This single-chain glycoprotein has a molecular weight of about 90,000, with 790 amino acids and 22 disulfide bonds that serve to maintain its convoluted three-dimensional structure (61,62). Plasminogen is converted to plasmin by a specific proteolytic cleavage yielding two chains, bound together by two disulfide linkages; this product is the catalytically active form of the protease, and the overall process is customarily denoted plasminogen activation.

Since the blood level of plasminogen is rarely a rate-limiting factor in its activation, thrombolytic therapy is ideally designed to activate this precursor at the site of thrombus decomposition. The thrombolytic agents currently being used clinically include streptokinase; anistreplase [81669-57-0], an anisoylated-plasminogen–streptokinase-activator-complex; urokinase; scuPA, [82657-92-9], ie, single-chain urokinase plasminogen activator (Saruplase, prourokinase); and tissue plasminogen activator, tPA (Alteplase).

Streptokinase. The fibrinolytic activity of streptokinase, isolated from strains of hemolytic *Streptococci*, was first demonstrated in 1933 (63). Streptokinase is a secreted protein product inasmuch as filtrates free of demonstrable bacteria were found to dissolve fibrin clots with rapidity. Streptokinase has a molecular weight of about 47,000 with a single chain of 415 amino acids; there are no intramolecular disulfide bonds (64). The complete nucleotide sequence of the gene encoding the RNA for this protein has been reported (65,66).

The fibrinolytic activity of streptokinase, mediated by generation of plasmin from plasminogen, is initiated by binding to the β-chain, ie, Val 561-Asp 790, of plasminogen (67,68). This results in a conformational change in the plasminogen molecule which, in turn, acts as a specific protease in activating a second plasminogen molecule to plasmin (69,70). To achieve optimal thrombolysis, the ratio of streptokinase to plasminogen should be approximately 1:10 (71).

Indications for treatment with streptokinase include acute occlusion of arteries, deep vein thrombosis, and pulmonary embolism. Streptokinase therapy in coronary thrombosis, which is the usual cause of myocardial infarction (54,71,72), has proved to be valuable. In this frequently fatal condition, the enzyme is administered intravenously at a dose of 1.5 million units over 60 min, or given by intracoronary infusion at a 20,000- to 50,000-unit bolus dose followed by 2000 to 4000 units min for 60 min; therapy must be instituted as soon as practicable after the diagnosis of heart attack is made. For deep vein thrombosis, pulmonary embolism, or arterial occlusion, streptokinase is infused at a loading dose of 250,000 units given over 30 min, followed by a maintenance dose of 100,000 units over a 60-min period.

Streptokinase has an initial plasma half-life ($t_{1/2\alpha}$) of 18 min, and a β half-life of 83 min (73); it is well recognized that the thrombolytic efficacy of the enzyme decreases as the age of the thrombus increases; thus, thrombolysis is significantly decreased when therapy is initiated more than three hours after an occlusion (74).

Several clinical trials have been conducted with streptokinase administered either intravenously or by direct infusion into a catheterized coronary artery. The results from 33 randomized trials conducted between 1959 and 1984 have been examined (75), and show a significant decrease in mortality rate (15.4% o) in enzyme-treated patients vs matched controls (19.2% o). These results correlate well with an Italian study encompassing 11,806 patients (76), in which the overall reduction in mortality was 19% o in the streptokinase-treated group, ie, 1.5 million units administered intravenously, compared with placebo-treated controls. The trial also shows that a delay in the initiation of treatment over six hours after the onset of symptoms nullifies any benefit from this type of thrombolytic therapy. Conversely, patients treated within one hour from the onset of symptoms had a remarkable decrease in mortality (47% o). The benefits of streptokinase therapy, especially in the latter group of patients, was still evident in a one-year follow-up (77). In addition to reducing mortality rate, there was an improvement in left ventricular function and a reduction in the size of infarction. Thus early treatment with streptokinase is essential.

In a more extensive international trial, 17,187 patients were treated intravenously with streptokinase alone, aspirin alone, a combination of streptokinase and aspirin, or placebo (78). Streptokinase and aspirin were equally effective in treating acute myocardial infarction, each decreasing mortality by 25% o; their combination further reduced mortality by 42% o. A significant reduction in mortality was seen even in those patients treated up to 24 hours after the onset of symptoms.

In addition to the intravenous route, streptokinase is also administered by the direct intracoronary route. In a Dutch study with 533 patients treated intracoronarily, a significant reduction (12% o) in mortality was noted (79). Of the two methods of administration, the intravenous route appears to be the method of choice because of ease of administration and the shorter lag period before therapy is initiated; the lag period observed in studies utilizing the intracoronary route is attributable to the time required for catheterization.

The thrombolytic efficacy of streptokinase treatment may be compromised by the presence of antibodies to the enzyme in the patient's blood. These neutralizing antibodies may arise because of a prior streptococcal infection, or prior streptokinase treatment (80–82). Titers of antibodies sufficient to neutralize a complete dose of 1.5 million units of streptokinase may be present even one year after enzyme treatment (83).

Since streptokinase has no clot-selective property, its therapeutic levels cause extensive activation of circulating plasminogen to plasmin. The increased plasmin levels not only cause degradation of the offending thrombus (therapeutic effect), but also cause degradation of coagulation factor V, factor VIII, and fibrinogen, and a reduction in α₂-antiplasmin (84). All of these factors produce a systemic lytic state that significantly increases the potential for hemorrhage, a serious hazard of thrombolytic therapy. Bleeding complications, including hemorrhagic strokes, have occurred, although infrequently, after administration of this enzyme. Another important side effect of therapy is related to the antigenic potential of streptokinase. Since the enzyme is obtained from bacterial sources, it can induce immune responses from allergy (fever, rash) to life-threatening anaphylactic shock (0.1% o in one study (76)). However, the benefits of fibrinolytic activity with this enzyme clearly outweigh such side effects as allergic reactions, hypotension, or minor bleeding. Moreover, in extensive clinical trials, serious side effects have proven to be rare; eg, there were no significant differences in the total number of

strokes in the treated vs the placebo groups.

Anistreplase (APSAC). The general drawbacks of streptokinase therapy are its short plasma half-life (18–20 min) and its poor clot selectivity. These drawbacks have been addressed by making conjugates of the enzyme. Anistreplase is a complex of streptokinase and acylated human Lys-plasminogen, synthesized by a noncovalent attachment of an acyl group to the active center of a streptokinase-Lys-plasminogen complex (85,86). Streptokinase on its own is inactive as a plasminogen activator, and must first form a complex with circulating plasminogen; this complex converts the second plasminogen molecule to plasmin. Anistreplase, on the other hand, is a fully active plasminogen activator. More importantly, it activates plasminogen at the site of deposition of the thrombus (85,86).

Anistreplase has a considerably longer α half-life than streptokinase, ie, 90 min compared to 20 min (87,88). Moreover, it does not require prolonged infusion to achieve its thrombolytic effects. Anistreplase was found to be highly effective after a single intravenous dose of 30 units over a 5-min period compared to a 60-min infusion of 1.5 million units of streptokinase (89–94). In direct comparative studies, anistreplase was as effective as intracoronary (95,96) and intravenously (96–100) administered streptokinase. In a randomized, double-blind, placebo-controlled study (AIMS trial) with 1004 patients given this modified enzyme, the 30-day mortality rate was 12.2° for patients receiving placebo, compared to 6.4° for patients who received 30 units of anistreplase intravenously within six hours of the onset of symptoms (101).

The side effects of anistreplase appear to be similar to those of streptokinase, including immune reactions and a systemic lytic state conducive to hemorrhage.

Urokinase. This trypsin-like serine protease consists of two polypeptide chains held together by a single disulfide bond. Urokinase exists in two molecular forms, ie, a high mol wt form of about 55,000, and a low mol wt variant of 33,000 (102,103). The half-life of urokinase, 10 min, is shorter than that of streptokinase (104). One of the more prominent disadvantages of streptokinase or anistreplase therapy is the antigenic potential; urokinase is a naturally occurring enzyme present in human urine, plasma, and tissues, and does not provoke an immune response directed against it. Urokinase can directly activate plasminogen, but is a nonspecific plasminogen activator, ie, it lacks clot selectivity, although the presence of fibrin does slightly stimulate the activation of plasminogen by urokinase (105). In early studies, urokinase was isolated from kL of human urine; in the early 1990s, recombinant DNA technology is used to manufacture the enzyme in cultured bacteria.

Compared to streptokinase, urokinase has been less extensively studied because of its high cost, ie, about 10 times that of a comparable treatment with streptokinase. In addition to the indications described for streptokinase, urokinase is indicated for use in patients with prior streptokinase treatment, or prior *Streptococcal* infection. Urokinase is commonly used at a loading dose of 4400 units kg, with a maintenance intravenous infusion dose of 4400 units kg h for thromboses other than acute myocardial infarction. In the latter case, a much larger dose, ie, 0.5–2.0 million units h or a bolus dose of 1.0 million units followed by a 60-min infusion with 1.0 million units, has been found optimal (106). An intracoronary dose of 2000 units min for two hours was used in one comparative study with intracoronary streptokinase (107). In this study, urokinase exhibited efficacy equivalent to streptokinase with fewer side effects. Other studies with intracoronary urokinase have administered doses ranging from 2,000 to 24,000 units min with a reperfusion efficacy of 60–89° (108–112). In another urokinase trial, 2.0 million units were administered intravenously, resulting in a thrombolytic efficacy of 60° (113). Effectiveness in terms of reduction in mortality rate has not been determined because of the small number of patients studied.

Overall, urokinase is better tolerated than streptokinase with no significant hypotension or allergic reactions. Intracoronary administration of urokinase, at a dose comparable in efficacy to streptokinase, also causes less reduction of fibrinogen, resulting in significantly fewer bleeding complications (107). It is generally accepted that, apart from the immune reactions associated with streptokinase therapy, the two enzymes are essentially the same as regards other side reactions.

Single-Chain Urokinase Plasminogen Activator (scuPA). This is the single-chain precursor of urokinase, with an enhanced fibrin specificity and clot selectivity (114–117). scuPA (pro-urokinase) was first purified from human urine using a fibrin-Celite column (118). By selective proteolysis, it activates only those plasminogen molecules directly associated with a fibrin clot. This results in a lag period until additional plasminogen molecules are exposed by the released plasmin at the clot surface (119). scuPA is a relatively new thrombolytic agent; its first clinical trial was reported in 1986 using a preparation derived from a cultured human renal adenocarcinoma cell line (120). Four of the six patients in this study showed beneficial therapeutic effects. This was followed by a multicenter trial in which patients were treated intravenously with 15–60 mg scuPA, urokinase, or its combination, ie, 200,000 units of bolus urokinase followed by a 60-min infusion with 48 mg scuPA (121–124). The patency rate, used as an index of reopening of a thrombosed vessel, was 55° for scuPA; that of the combination was 82°. Inclusion of urokinase in the regimen increases the patency rate without significantly changing the fibrinogen, plasminogen, or α 2-antiplasmin degradation profile observed with scuPA alone. Pretreatment with urokinase reduces the lag period of the first phase of scuPA action, and therefore acts synergistically with it. Although several studies have demonstrated the fibrinolytic specificity of scuPA, the overall side effects of this agent in clinical trials are quite similar in nature to those of its nonspecific analogue, urokinase.

Tissue Plasminogen Activator (tPA). While streptokinase and urokinase can effectively induce clot dissolution in the majority of patients if given early, they lack clot specificity. Treatment with these enzymes results in a systemic lytic state attributable to their degradative action on circulating fibrinogen. Tissue plasminogen activator (tPA) was developed to achieve rapid and specific thrombolysis.

tPA was first identified in 1966 as a naturally occurring serine protease (125). However, it was not characterized until 1979 when a sufficient quantity of tPA was purified from human uterine tissue (126). A more extensive characterization of tPA was performed following its purification from a human melanoma cell line that produces large quantities of the molecule (127,128). Subsequent to these studies, human tPA was cloned and purified by recombinant DNA technology (129). The molecular weight of tPA is about 70,000, with 527 amino acids in its single-chain form. The single-chain molecule is converted to its active two-chain form by hydrolysis of the Arg275–Ile276 peptide bond (130). At the clot surface, single-chain tPA generates small quantities of plasmin by plasminogen activation; these then initiate the thrombolytic cycle by generating additional plasmin at the clot surface, resulting in dissolution. The presence of fibrin dramatically increases the catalytic activity of tPA (131,132). The initial half-life of exogenous tPA in humans is about 5 to 8 min (133).

In the first clinical trial of tPA for acute myocardial infarction, the molecule purified from melanoma cell cultures was used (134). A reperfusion rate of 85°, ie, six of seven patients, was observed after infusion of 6 to 14 mg of tPA at the rate of 12–24 mg h. More importantly, no changes in fibrinogen, α 2-antiplasmin, or circulating plasminogen levels were detected following treatment. With the availability of human tPA produced by recombinant DNA methods, additional clinical studies have been conducted. The *in vitro* thrombolytic activity of recombinant tPA and of tPA obtained from melanoma cells was identical (135). In clinical trials with recombinant tPA for coronary thrombosis, a 73° reperfusion rate was observed after the intravenous infusion of 0.5 mg kg for 60 min (136); an 87° rate was observed when the dose was raised to 0.75 mg kg given over 120 min (137). The mean time to recanalization averaged 46 $\pm$ 18 min. These trials show that recombinant tPA is efficacious, and that its use is accompanied by minimal side effects. Another clinical trial, designed to compare the efficacy of recombinant tPA with streptokinase, yielded a 70° patency rate of the occluded artery in the recombinant tPA group, given 0.75 mg tPA kg over 90 min, compared to a 55° patency rate in the streptokinase group, given 1.5 million units of enzyme over 60 min (138). Although there was no significant difference in the percentage patency achieved in these studies, there was a significantly greater degradation of fibrinogen in the streptokinase treated group. Thus, fibrinogen levels fell below 0.5 g L in 90° of the streptokinase treated patients, but in only 5° of the recombinant tPA patients. In another comparative trial, 80 mg recombinant tPA over three hours was compared to 1.5 million units of streptokinase. The recombinant tPA treatment resulted in a 60° reperfusion rate at 90 min vs a 35° rate when streptokinase was used alone (137). Additional randomized clinical trials show recombinant tPA to be more efficacious than streptokinase in coronary thrombolysis, with an average of 75° patency for recombinant tPA and about 50° with streptokinase (138–143). In one controlled, randomized study with 5013 patients, a significant decrease in mortality was observed with recombinant tPA compared to patients given placebo (144). However, in two more recent (ca 1990) large-scale trials no significant differences between streptokinase and the tPA treatments were observed: in a study of 12,490 patients (145), in-hospital mortality and severe left ventricular damage were used as end points; in the International Study Group's SK tPA trial in which 8401 patients were randomly assigned, mortality was used as the sole end point. In a more recent and larger ISIS-3 trial (147) comprising 46,092 patients, direct comparisons were made between the three thrombolytic agents, streptokinase, recombinant tPA, and APSAC; no significant differences among the three agents could be demonstrated. A 35-day mortality rate was used as the primary end point of this trial, and the mortality rates were streptokinase, 10.5°; recombinant tPA, 10.3°; and APSAC, 10.6°. In addition, patients

in the streptokinase group had a significantly lower incidence of strokes than patients in the recombinant tPA and the APSAC groups. This is paradoxical since the development of recombinant tPA was based on the premise that it is clot-specific and thus should not have produced bleeding complications.

Combination of Plasminogen Activators. Since each of the five agents discussed achieve thrombolytic activity in a unique fashion on the basis of individual molecular structures, various combinations of these agents are being evaluated with the goal of minimizing side effects, especially the risk of hemorrhage, and of augmenting efficacy (148). For example, administration of a small amount of urokinase prior to treatment with scuPA results in an increase in patency rate with no further decrease in circulating fibrinogen (122,123). The most extensively studied combination is of scuPA and tPA. Synergistic thrombolytic activity of this combination has been shown both *in vitro* (149–151), and *in vivo* (152,153). In an initial clinical study with nine patients, combination of low doses of scuPA (10 mg) and tPA (10 mg) infused over a 60-min period resulted in a patency rate of 78° (102). In a slightly larger patient group (38 patients), the combination of recombinant tPA (12 mg) and scuPA (48 mg) infused over 30- and 40-min, respectively, resulted in a patency rate of 61° at 60 min and 82° at 90 min, with minimal side effects (154). Similar results have been obtained in two other human trials (155,156). Combination studies have also been designed to reduce the amount of tPA used in these treatments because of its high cost. Combination of one-half the usual dose of recombinant tPA, ie, 50 mg instead of 100 mg, with a full dose of streptokinase (1.5 million units) resulted in a 75° patency rate (157); this rate is comparable to that achieved in other clinical trials using single thrombolytic agents.

One drawback of thrombolytic therapy is a high incidence of reocclusion. In a report using a canine model, inclusion of heparin [9005-49-6/ (anticoagulant therapy) in the treatment prevented this side effect (158). The combination of aspirin [50-78-2/ (antiplatelet therapy) and streptokinase (thrombolytic therapy) has also shown significant therapeutic advantages (78). Although additional work is needed to establish the thrombolytic advantage of various combinations, preliminary results in this area indicate promise in terms of increased efficacy and reduced side effects.

Replacement Therapy for Inherited Enzyme Deficiencies. Correction of inborn errors of metabolism is one of the principal goals of enzyme treatment. For example, lysosomal storage diseases arise from a genetically determined deficiency in the activity of a single specific enzyme, resulting in the accumulation of substrates in the lysosomes of affected cells; this accumulation in turn results in cellular engorgement, and death. Kinetic studies predict that clinical disease will not occur until enzyme activity falls below 10° of normal (159,160). Well-known examples of disorders in this group of thesauroses, ie, storage disorders, are TaySachs disease (hexosaminidase deficiency), Gaucher's disease (glucocerebrosidase deficiency), Krabbe's disease (galactocerebrosidase deficiency), the mucopolysaccharidoses (various eponyms), and mannosidosis (159). With the recognition that many of these disorders are the result of inadequacies of lysosomal enzymatic catabolism, it was anticipated that administration of the deficient enzyme might be used to dispose of the respective pathologic accumulations. Such treatment has yielded symptomatic improvement in select cases, but the task of forcing exogenous enzymes to enter the cytoplasm or nucleus of affected cells has not yet been successfully addressed.

One of the principal problems in cystic fibrosis is malnutrition (161); because malnutrition affects growth, and contributes to the severity of pulmonary involvement, it plays a decisive role in shortening the survival of subjects with this condition. Fat malabsorption in these patients has been attributed to pancreatic insufficiency. The use of pancreatic enzymes in these patients helps in the correction of malabsorption, thereby restoring a more normal pattern of growth (162). Pancreatin [8049-47-6/], cotazym capsules, or Viokase tablets, alone or in combination with bicarbonate and or cimetidine to reduce gastric acidity, administered to patients with cystic fibrosis result in increased fat absorption, normal levels of lipid-soluble vitamins, and improved growth in ~70% of the patients given these regimen (163–165). However, the poor dose-response relationship between different preparations (166) and the correction of malabsorption is peculiarly weak. A pancreatic preparation (Creon) having an improved lipase trypsin ratio was tested in 15 patients with severe pancreatic insufficiency; this formulation produced a dose-related increase in fat absorption as well as in weight gain (167). The combination of pancreatin and calcium gluconate (168) has been shown to be useful in reducing steatorrhea resulting from chronic pancreatitis. Enteric-coated preparations sensitive to pH can be used to ensure that a given enzyme is released only in the alkaline medium of the duodenum or proximal jejunum (169). Microsphere preparations, consisting of enteric-coated spheres of lipase and protease inside a gelatin capsule, were superior to conventional enzyme therapy for treating fat malabsorption in patients with cystic fibrosis (170–173). In these patients, such microsphere preparations were able to reduce both steatorrhea and symptoms such as abdominal pain and frequency of evacuation (170–172).

Adenosine deaminase [9026-93-1/ (ADA) deficiency in children causes profound immunodeficiency, analogous to that seen in AIDS, leading to frequent infections. This condition has been successfully treated in a few subjects with poly(ethylene glycol)-modified bovine ADA (173,174). These patients had 1° or less of ADA activity in their erythrocytes and circulating mononuclear cells. Following treatment, the patients showed normal circulating T-lymphocytes and marked improvement in weight gain as well as in the incidence of infection.

Lactose [63-42-3/ intolerance is a genetic deficiency of lactase [9031-11-2/ which is responsible for hydrolyzing lactose to form D-glucose [50-99-7/ and D-galactose [59-23-4/]. In affected infants, ingestion of lactose, primarily from milk, causes diarrhea, cramping, bloating, and abdominal pain (175). It has been demonstrated that the addition of β-galactosidase to milk at mealtime hydrolyzes lactose effectively and reduces lactose malabsorption and intolerance to milk (175,176). β-Galactosidase from *Aspergillus oryzae* (Lactrase) administered along with milk resulted in temporary reversal of lactose malabsorption in five of nine recipients (176).

Alglucerase [9001-22-3/ (β-glucocerebrosidase, β-glucosidase, ceredase), produced by Genentech, was approved by the FDA for the treatment of Gaucher's disease (177). This enzyme is obtained from human placenta. Gaucher's disease is a functional deficiency of β-glucocerebrosidase in tissue macrophages which become engorged, and thus are termed Gaucher's cells; they are found in the liver, spleen, and bone marrow. This deficiency leads to hepatomegaly. Skeletal complications including osteoporosis and osteonecrosis with secondary pathological fractures are a common feature of Gaucher's disease.

Antihemophilic factor [9001-28-9/ (AHF) is a protein found in normal plasma that is necessary for clot formation. It is needed for transformation of prothrombin to thrombin. Administration of AHF by injection or infusion can temporarily correct the coagulation defect present in patients with hemophilia. Antihemophilic factor VIII (Alpha Therapeutic) has been approved by the FDA as replacement therapy in patients with hemophilia B to prevent bleeding episodes, and also during surgery to correct defective hemostasis (178).

Depolymerizing Enzymes. Depolymerizing enzymes belong to a special class of hydrolases which cleave peptide bonds using water as co-substrate. They are used widely by the topical route as antiinflammatory agents. They reduce inflammation and edema by digesting and dissolving the proteinaceous debris found in inflammatory exudates (179).

Another subclass of proteases attacks internal peptide bonds and liberates large peptide fragments. Bromelain, a plant protease derived from the stem of the pineapple plant, can even produce detectable serum proteolysis after oral administration (180). Oral therapy with bromelain significantly reduces bruising that stems from obstetrical manipulations (181). Bromelain–pancreatin combinations have been more effective in digestive insufficiency compared to either pancreatin or placebo (182,183). Bromelain may also enhance the activity of antibiotics, especially tetracycline, when administered concurrently (184).

Chymopapain [9001-90-6/], derived from the latex of the papaya tree, produces improvement in lower back pain and sciatica in the majority (75°) of recipients (185–189) when injected into the lumbar intervertebral disks of patients suffering from herniated disk (the nucleus pulposus). This treatment degrades the proteoglycans of the diseased nucleus pulposus, resulting in shrinkage of the disk and reduction of pressure on the nerve roots (190).

Collagenase specifically catalyzes the hydrolysis of collagen, and is used in debridement of dermal ulcers and burns (191). It, like chymopapain, is also useful in the treatment of herniated lumbar disks (192,193). The rationale for collagenase treatment in this instance is based on the preponderance of collagen in herniated disk tissue, and the inability of other enzymes to dissolve collagen (194).

Papain has been shown to produce marked reductions of obstetrical inflammation and swelling, and of edema resulting from dental surgery (195–197). In a controlled experiment in dogs with gunshot wounds, a neutral protease of bacterial origin was shown to proteolyze devitalized tissue, cleaning the wounds, and promoting granulation and epithelialization (198). Streptokinase–streptodornase [37340-82-2/], containing both protease and nuclease activities, has long been used in the treatment of inflammatory exudates. The streptokinase component hydrolyzes plasminogen resulting in the lysis of fibrin clots, while the streptodornase component depolymerizes deoxyribonucleic acid (DNA), leading ultimately to liquefaction of exudates. This combination also was found to be useful in the treatment of empyema, hemothorax, and loculated or clotted pleural effusions (199,200). Streptokinase also has been used in patients maintained on continuous peritoneal dialysis who have developed recurrent bacterial peritonitis resistant to therapy. Bacterial sequestration within fibrin clots located on indwelling catheters contributes to the resistance of this type of infection to standard antibiotic therapy.

Streptokinase dissolves such fibrin clots, thereby permitting effective antibiotic therapy (201).

The use of fibrin [9001-31-4] to seal wounds in ophthalmic surgery was first demonstrated in 1945 (202). Fibrin has also been used along with conventional sutures as a sealant in lamellar keratoplasty (203). Such fibrin glue effectively reduces pulmonary air leakage in animals and humans with ruptured lungs, and can also be utilized for adhesion of prematurely ruptured membranes during pregnancy (204–207). Other applications include repair of traumatically ruptured spleen in children (208), sealing of chronic gastric ulcers in patients with repeated episodes of bleeding (209), repair of leakage of lymph from a femoral cutdown site in a neonate (210), rejoining of severed Achilles' tendon (211), and repair of delicate peripheral nerves in microneurosurgery (212).

Deoxyribonuclease (DNAase), an enzyme that degrades deoxyribonucleic acid, has been used in patients with chronic bronchitis, and found to produce favorable responses presumably by degrading the DNA, contributed by cell nuclei, to inflammatory mucus (213). Lysozyme [9001-63-2] hydrolyzes the mucopeptides of bacterial cell walls. Accordingly, it has been used as an antibacterial agent, usually in combination with standard antibiotics. Topical applications are also useful in the debridement of serious burns, cellulitis, and dermal ulceration.

The proteolytic enzymes, trypsin, chymotrypsin, and chymoral [8076-22-0], in combination, have been used for the treatment of post-operative hand trauma, athletic injuries, and sciatica (214–216). Trypsin has also been used successfully in treating hyaline membrane disease of newborn babies, a condition usually fatal without treatment (217). Immobilized preparations of trypsin are useful in treating acute radiation cystitis following pelvic x-irradiation therapy (218).

Superoxide dismutase has been approved by the FDA for preventing reperfusion injury or damage to donor organ tissue (178). This enzyme is prepared by recombinant DNA technology and marketed by Bristol-Myers and Pharmacia-Chiron.

Improvements of the electrocardiograms of patients with acute myocardial infarction have been demonstrated following hyaluronidase treatment (219). Hyaluronidase destroys hyaluronic acid, a mucopolysaccharide, thus allowing vital molecules to penetrate the normally impermeable connective tissue barrier. Accordingly, hyaluronidase is believed to limit the extension of a myocardial infarct by increasing the diffusion of nutrients through the liquefied ground substance. Hyaluronidase also has been used to reduce intraocular pressure during cataract surgery (220,221). Intrathecal treatment with hyaluronidase has been used in pediatric practice for the management of hydrocephalus resulting from tubercular meningitis (222). In extravasation injuries of children, hyaluronidase, injected after an intravenous infiltration, significantly reduces tissue damage (223).

Enzymes as Antidotes. Rhodanese [9026-04-4] given along with thiosulfate to counteract cyanide poisoning in mice (224) was the first enzyme used as an antidote. This combination raised the LD₅₀ of potassium cyanide in mice by eightfold (224).

Superoxide dismutase, an enzyme that decomposes the highly toxic oxygen free radical O₂, has been put into veterinary use as an antiinflammatory agent with efficacy in the treatment of traumatic arthritis of horses (225). The identification of free-radical toxicity as being operative in poisonings with an increasing variety of pharmacologic agents is expected to increase the use of superoxide dismutase (225,226). Superoxide dismutase has also been shown to exert antiinflammatory effects on uveitis (227–230), as has catalase, and to protect murine and human bone marrow progenitor cells from the effects of radiation (231). Liposomal superoxide dismutase has been used to treat radiofibrosis caused by excessive radiation and leading to fibrosis at a local target site (232). Superoxide dismutase prepared by recombinant technology prevents reperfusion injury of the ischemic spinal cord in dogs (233).

Glutathione peroxidase [9013-66-5] oxidizes glutathione, and helps to remove inorganic and organic hydroperoxides (227). It exhibits antiinflammatory activity in experimental uveitis of rats (234).

Bilirubin oxidase [80619-01-8], derived from *Myrothecium verrucaria*, was modified with polyethyleneglycol; when this conjugate was injected intravenously to jaundiced rats, the plasma bilirubin dropped to normal levels. This approach might have potential in the treatment of hyperbilirubinemia, fulminant hepatitis, and neonatal bilirubin encephalopathy (177).

In some patients with IgA nephropathy (IgAN), intraglomerular coagulation plays a role in depositing fibrinogen (235,236). IgAN patients treated with urokinase show a marked improvement in urinary protein concentration, serum creatinine, and blood urea nitrogen levels (237).

Immobilized, Derivatized, and Entrapped Enzymes

Immobilized Enzymes. Immobilized catalase was shown in the early 1970s to correct the enzyme deficit in genetically acatalasemic mice (238). The ideal goals of enzyme modification are to increase enzyme stability *in vivo* by rendering them resistant to the action of endogenous proteinases, inhibitors, and antibiotics; increase their stability during storage; decrease their immunogenicity, antigenicity, and toxicity; decrease the dose of a given enzyme preparation; and prolong enzyme action as indicated by the enzyme's pharmacokinetic profile (239–241). With the development of techniques for binding enzymes to insoluble supports, immobilized enzymes have been used not only in clinical analysis but also for therapeutic purposes. For example, L-asparaginase therapy with an extra-corporeal device in patients with malignant lymphoma has resulted in repeated remissions of metastases (242). This technique utilizes L-asparaginase covalently bound to the surface of a battery of plates contained in a portable chamber which has been inserted between an artery and vein.

Enzyme Conjugates. One approach used to prolong residence time of a given enzyme in the circulation is to conjugate that enzyme with albumin [103218-45-7], a natural plasma protein. Cross-linked preparations of albumin and either uricase [9002-12-4], which oxidizes urea to form allantoin [97-59-6], or L-asparaginase, using the bifunctional reagent glutaraldehyde, were first described in 1974 (243). Conjugates of albumin with L-asparaginase or α -1,4-glucosidase were shown to be 4 to 10 times more stable to denaturation by either heat or trypsin than the native enzyme (244). Albumin conjugates of α -1,4-glucosidase, superoxide dismutase, and uricase were also shown to have notably longer plasma halflives than their unconjugated counterparts (245).

One limitation of enzyme replacement therapy is the targeting of enzyme proteins to appropriate sites of substrate accumulation. Administration of a cholesterol esterase conjugated to albumin results in the degradation of pathologic cholesterol ester accumulations within the lysosomes of fibroblasts from a patient with cholesterol ester storage disease (246).

Hemodialysis with microencapsulated urease and an ammonia ion adsorbent, zirconium phosphate [13772-29-7], has been used (247) to delay the onset of dialysis therapy in patients retaining some renal function, and to reduce the time between dialysis treatment.

Poly(ethylene glycol) (PEG) molecules attached to adenosine deaminase (ADA) have been used in patients exhibiting symptoms of the severe combined immunodeficiency syndrome (SCID) caused by ADA deficiency. The modified enzyme has a plasma half-life of weeks as compared to the unmodified enzyme (minutes) (248). PEG-L-asparaginase has induced remissions in patients with non-Hodgkin's lymphoma (248). However, one disadvantage of PEG-enzyme treatment is its expense, ie, a year's treatment costs about \$60,000 (248).

Erythrocyte Entrapment of Enzymes. Erythrocytes have been used as carriers for therapeutic enzymes in the treatment of inborn errors (249). Exogenous enzymes encapsulated in erythrocytes may be useful both for delivery of a given enzyme to the site of its intended function and for the degradation of pathologically elevated, diffusible substances in the plasma. In the use of this approach, it is important to determine that the enzyme is completely internalized without adsorption to the erythrocyte membrane. Since exposed protein on the erythrocyte surface may elicit an immune response following repeated sensitization with enzyme loaded erythrocytes, an immunologic assessment of each potential system in animal models is required prior to human trials (250).

Chemically or enzymatically modified preparations of enzymes present another potential immunologic problem since modifications designed to stabilize enzyme activity may produce new antigenic determinants. For example, the attachment of carbohydrate moieties to enzymes using glycopeptides, carbohydrate polymers, or individual sugars may alter the *in vivo* fate of a given enzyme, causing it to become immunogenic (250). Moreover, addition of carbohydrate residues may also expose new antigenic sites by inducing conformational changes in the protein; additionally, the carbohydrate moiety itself might act as a hapten or, if large enough, as a neo-antigen in its own right (250).

As of this writing there are eight methods of erythrocyte entrapment. Six methods depend on loading via hypotonic exchange, one method depends on chlorpromazine-induced endocytosis, and one depends on voltage-step induced transitory permeation. Although only minor leakage of entrapped enzyme is detected by these methods, up to 11% of entrapped glucose is released within three hours of entrapment (251).

Erythrocyte-entrapped β -glucuronidase has been retained in the circulation and in the liver of β -glucuronidase-deficient mice for longer periods of

time than intravenously administered free enzyme (251). Entrapment in autologous erythrocytes may provide an effective means of optimizing the delivery and protection of exogenous enzymes in the treatment of selective lysosomal storage diseases, such as Fabry disease and Type I Gaucher disease.

Economic Aspects

There has been a steady growth in the economic importance of therapeutic enzymes, with sales reaching hundreds of millions of dollars per year as of 1992. Table 2 lists the trade names and costs of some of the more commonly prescribed enzyme preparations. Despite the magnitude of use of these products, the manufacture and sale of therapeutic enzymes represents a comparatively small fraction of the production and profits of the pharmaceutical houses that market them.

Table 2. Economic Importance of Therapeutic Enzymes.

Trade name	Cost, ^a \$	Contents	Manufacturer
<i>Thrombolytic enzymes</i>			
Activase	550 20 mg	lyophilized recombinant tPA	Genentech
Abbokinase	263 250,000 IU	lyophilized urokinase	Abbott
Kabikinase	600 250,000 IU	streptokinase	SKF
Streptase	157 250,000 IU	streptokinase	Hoechst-Roussel
Eminase		anisoylated plasminogen	Beecham
<i>Oncolytic enzymes</i>			
Elspar	40 10,000 IU	L-asparaginase	Merck
<i>Ophthalmic enzymes</i>			
Catarase	22 150 IU	chymotrypsin	Lolab
Zolyse	28 750 IU	chymotrypsin	Alcon
<i>Topical enzymes^b</i>			
Travase	34 14 g	sutilains	Boots
Santyl	28 15 g	collagenase	Knoll
Elastase	11 10 g	fibrinolysin deoxyribo-nuclease	Fugisawa
Granulex ^c	2.10 10 g	trypsin papain	Hickam
Panifil ^d	2.30 10 g	papain urea	Rystan
Panifil White ^e	5.60 10 g	papain urea	Rystan
<i>Digestive enzymes^f</i>			
Entozyme	18 100	pancreatin pepsin bile salts	Robins
Viokase	15 100	pancreatin pepsin bile salts	Robins
Pancrease	30 100	pancreatin pepsin bile salts	McNeil
Cotazyme-S	24 100	pancreatin pepsin bile salts	Organon
Festal II	17 100	pancreatin pepsin bile salts	Hoechst-Roussel
Creon	32 100	pancreatin pepsin bile salts	Reid-Rowell
Ilozyme	77 250	pancreatin pepsin bile salts	Adria
Hi-Vegi-Lip	15 100	pancreatin lipase protease amylase	Freeda
Dizymes	9 100	pancreatin lipase protease amylase	Recser Labs
Digestozyme	8 100	pancreatin pepsin HCl	Various
Acro-Lase	5 50	amylase protease lipase cellulose	Arco
Enzobile	13 100	pancreatic enzyme cellulase pepsin	Mallard
Kanulase	14 50	amylase protease lipase pepsin bile extract cellulase glutamic acid	Sandoz
Digestalin	3 100	pancreatin pepsin papain activated charcoal bismuth barberis hydrastis	Vortech
Phazyme-PB	15 100	amylase protease lipase simethicone phenobarbital	Reed and Camrick
Kutrase	29 100	amylase pancrease lipase cellulase hyoscyamine phenyltoloxamine	Schwarz
Donnazyme	20 100	pancreatin pepsin homatropine hyoscyamine atropine scopolamine phenobarbital	Robins
Panase	21 100	pancreatic enzymes	Qualitest

^a Costs (per units indicated) obtained from Ref. 252.
^b Used to selectively digest necrotic tissue in wounds and burns.
^c 12 2 oz.
^d 103 1 lb.
^e 16 1 oz.
^f Available in tablet or capsule formulations.

Health and Safety Factors

Repetitive doses of foreign proteins may produce severe immunologic reactions, ranging from mild allergy to anaphylactic shock and death. Parenteral administration of metabolically active enzymes can be acutely toxic on account of their biochemical effects (253,254). However, large metabolically effective doses of these products may be less toxic than small doses, due to immune paralysis, wherein large doses of antigen repress the expression of the complementary antibody. It is essential to circumvent such allergic responses in the enzyme therapy of genetic diseases, because replacement enzyme therapy may be required over a lifetime. By contrast, treatment of cancer with enzymes may be somewhat less problematic because of the attenuation of the immunologic reactivity of patients in many cases. A significant effort is being directed toward the use of human sources of enzymes, or of human-type enzymes produced in cultures of prokaryotic or eukaryotic organisms, with a goal of reducing or eliminating the problems of antigenicity.

For purposes of dosage, the specific activity of an enzyme is usually expressed as International Units (IU) rather than in terms of weight. However, unit measurements do not provide information on the absolute purity of a given product. Moreover, purity is not as critical an attribute for oral enzymes, as opposed to those administered parenterally, inasmuch as the gastrointestinal tract is capable of disposing of most inert contaminants.

For enzymes intended for parenteral use, the manufacturer must assure that the enzyme preparation is essentially pure and free of endotoxins. Electrophoretic and immunologic tests provide the requisite evidence of purity and homogeneity. Most importantly, the manufacturer must remove toxic impurities, eg, bacterial lipopolysaccharide (endotoxins) which might cause severe toxic reactions such as anaphylactic shock, fever, and vascular collapse.

All preparations of enzymes intended for parenteral use are tested for safety in lower animals under the conditions anticipated in clinical trials; ie, their use must be nonpyrogenic in the USP rabbit assay (255), and must be sterile. Such toxicologic studies are usually a prerequisite for approval by the FDA for the sale of such pharmaceuticals.

Since many therapeutic enzymes are still derived from bacterial sources, FDA requirements can serve to make the commercial preparations more expensive. However, toxicological examination of each lot may not be necessary when the purification procedures yield reproducible preparations.

BIBLIOGRAPHY

"Enzymes Therapeutic" in *ECT* 3rd ed., Vol. 9, pp. 225–240, by D. A. Cooney, G. Stergis, and H. N. Jayaram, Division of Cancer Treatment, National Cancer Institute, National Institutes of Health.

1. R. Emmerich, O. Low, and A. Korschun, *Cent. Baket.* **31**, 1 (1902).
2. O. T. Avery and R. D. Dubos, *J. Exp. Med.* **58**, 73 (1931); K. Goodner, R. Dubos, and O. T. Avery, *J. Exp. Med.* **55**, 393 (1931).
3. W. S. Tillett and R. L. Garner, *J. Exp. Med.* **58**, 485 (1933).
4. J. D. Broome, *Nature* **191**, 1114 (1961).
5. J. D. Broome, *Trans. N.Y. Acad. Sci.* **30**, 690 (1968).
6. A. P. Fletcher and co-workers, *J. Lab. Clin. Med.* **65**, 713 (1965).
7. E. Hug and W. K. Schubert, *J. Cell. Biol.* **35**, 1 (1967).
8. J. W. Marsh, J. Denis, and J. C. Wriston, *J. Biol. Chem.* **252**, 7678 (1977).
9. J. W. Baynes and F. Wold, *J. Biol. Chem.* **251**, 6016 (1976).
10. G. Gregoriadis, *Methods Enzymol.* **44**, 698 (1976).
11. Y. Fishman and N. Citri, *FEBS Lett.* **60**, 17 (1975).
12. K. F. Driscoll and co-workers, *J. Pharmacol. Exp. Ther.* **195**, 382 (1975).
13. W. T. Shier, J. T. Trotter, and D. T. Astudillo, *Int. J. Cancer* **18**, 672 (1976).
14. *Enzyme Nomenclature*, Academic Press, Inc., New York, 1984.
15. J. R. Uren and H. Lazarus, *Proc. Am. Assoc. Cancer Res.* **16**, 144 (1975).
16. J. C. Wriston and T. O. Yellin, *Adv. Enzymol.* **39**, 185 (1973).
17. P. H. Sato, *Drug Metab. Dispos.* **12**, 1 (1984).
18. E. Payr, *Muench. Med. Wochenschr.* **69**, 1330 (1922).
19. M. P. Gallagher, R. D. Marshall, and R. Wilson, *Essays Biochem.* **24**, 1 (1989).
20. H. Wade and co-workers, *Ann. N.Y. Acad. Sci.* **613**, 95 (1990).
21. H. B. Muss and co-workers, *Invest. Drugs* **8**, 125 (1990).
22. H. Jurgens and co-workers, *Klin. Padiatr.* **200**, 184 (1988).
23. B. Horowitz and co-workers, *Science* **160**, 533 (1968).
24. M. M. Hudson and co-workers, *Cancer* **65**, 2615 (1990).
25. C. Evans and co-workers, *Cancer* **65**, 2624 (1990).
26. J. S. Holcenberg in J. S. Holcenberg and J. Roberts, eds., *Enzymes as Drugs*, Wiley-Interscience, New York, 1981, p. 22.
27. J. S. Holcenberg and co-workers, *Cancer Res.* **39**, 3145 (1979).
28. A. S. D. Spiers and H. E. Wade, *Br. Med. J.* **29**, 1317 (1976).
29. M.-C. Wu and co-workers, *In Vitro* **18**, 750 (1982).
30. R. P. Warren and co-workers, *Cancer Res.* **40**, 4546 (1980).
31. J. C. Chuang, C. L. Yu, and S. R. Wang, *Clin. Exp. Immunol.* **82**, 469 (1990).
32. H. J. Rosenfield, K. A. Watanabe, and J. Roberts, *J. Biol. Chem.* **12**, 1011 (1976).
33. W. Kreis, *Cancer Treat. Rep.* **63**, 1069 (1979).
34. K. A. Carr and M. J. Arduino, *Bios* **52**, 13 (1981).
35. C. W. Abell, W. J. Stith, and D. S. Hodgins, *Cancer Res.* **32**, 285 (1972).
36. C. W. Abell, D. S. Hodgins, and W. J. Stith, *Cancer Res.* **33**, 2529 (1973).
37. V. Yu and co-workers, *Byull. Ekep. Biol. Med.* **109**, 458 (1990).
38. R. Issacs, *Fed. Proc.* **6**, 394 (1947).
39. M. D. Prager in R. L. Clark and co-eds., *Oncology 1970*, Vol. 2, Yearbook Medical Publishers, Inc., Chicago, Ill., 1971, p. 237.
40. R. S. Greenfield and D. Wellner, *Cancer Res.* **37**, 2523 (1977).
41. J. Roberts, *Proc. Am. Assoc. Cancer Res.* **18**, 31 (1977).
42. J. R. Uren and H. Lazarus, *Proc. Am. Assoc. Cancer Res.* **16**, 144 (1975).
43. J. Everse and co-workers, *Proc. Natl. Acad. Sci. USA* **74**, 472 (1977).
44. K. K. Kalghatgi and J. R. Bertino in Ref. 26, p. 77.
45. C. C. Levy and T. P. Karpetsky in Ref. 26, p. 103.
46. R. Piccoli and co-workers, *Ital. J. Biochem.* **39**, 242 (1990).
47. N. P. Zelenkova and co-workers, *Tsitol. Genet.* **22**, 31 (1988).
48. C. L. Jelsema and co-workers in Ref. 26, p. 259.
49. R. L. Simmons and co-workers, *Surgery* **70**, 38 (1971).
50. G. Sava and co-workers, *Anticancer Res.* **6**, 183 (1986).
51. G. Sava, L. Perissian, and S. Zorzet, *Clin. Expl. Metastasis* **6**, 245 (1988).
52. G. Sava, V. Ceschia, and G. Zabucchi, *Eur. J. Cancer Clin. Oncol.* **24**, 1737 (1988).
53. H. Maeda, Y. Matsumura, and A. Molla, *Cancer Res.* **47**, 563 (1987).
54. M. A. DeWood and co-workers, *Circulation* **68** (Suppl. 1), 39 (1983).
55. G. Baumgartner, *Wein Klin. Wochenschr. Suppl.* **174**, 3 (1987).
56. H. T. Abelson and co-workers, *Cancer Treat. Rep.* **62**, 629 (1978).
57. H. H. Sedlacek and co-workers, *Int. J. Immunopharmac.* **9**, 841 (1987).
58. J. G. Bekesi and J. F. Holland, *Recent Results Cancer Res.* **80**, 42 (1982).
59. B. R. Olin, ed., *Facts and Comparisons*, J. B. Lippencott Co., New York, 1992, p. 765.
60. F. Dukert, *Semin. Thromb. Hemostasis* **10**, 87 (1984).
61. K. C. Robbins and co-workers, *J. Biol. Chem.* **240**, 541 (1965).
62. D. Raum and co-workers, *Science* **208**, 1036 (1980).
63. W. S. Tillett and R. L. Garner, *J. Exp. Med.* **58**, 485 (1933).
64. E. C. De Renzo and co-workers, *J. Biol. Chem.* **242**, 533 (1967).
65. K. W. Jackson and J. Tang, *Biochemistry* **21**, 6620 (1982).
66. H. Malke, B. Roe, and J. J. Ferretti, *Gene* **34**, 357 (1985).
67. C. M. Ling, L. Summaria, and K. C. Robbins, *J. Biol. Chem.* **240**, 4213 (1965).

68. K. C. Robbins and co-workers, *Thrombos Haemostas* **50**, 787 (1983).

69. N. Heinburger in H. Kreuzer, ed., *Streptokinase-Behandlung der Thromboembolien und des akuten Myokardinfarktes*, Schafftauer Verlag, Stuttgart, Germany, 1982, p. 23.

70. M. Martin, *Thrombos. Daiibes. Haemorrh.* **33**, 586 (1975).

71. M. A. De Wood, *N. Engl. J. Med.* **303**, 897 (1980).

72. K. P. Rentrop, *N. Engl. J. Med.* **311**, 1457 (1984).

73. A. P. Fletcher, N. Alkjaersig, and S. Sherry, *J. Clin. Invest.* **37**, 1306 (1959).

74. D. O. Williams in E. J. Topol, ed., *Acute Coronary Intervention*, Alan R. Liss, New York, 1987, p. 137.

75. S. Yusuf and co-workers, *Eur. Heart J.* **6**, 556 (1985).

76. Gruppo Italiano per lo Studio della Streptochinasi Nell'Infarcto Miocardio (GISSI), *Lancet* **i**, 397 (1986).

77. A. P. Maggioni in Ref. 74, p. 119.

78. "Second International Study of Infarct Survival (ISIS-2)," *Lancet* **ii**, 349 (1988).

79. M. L. Simoons, *Lancet* **ii**, 578 (1985).

80. H. D. White and co-workers, *Brit. Heart J.* **64**, 177–181 (1990).

81. D. E. Vaughan and co-workers, *Circulation* **84**, 84–91 (1991).

82. M. Lynch and co-workers, *Brit. Heart J.* **66**, 139–142 (1991).

83. D. Massel and co-workers, *Circulation* **82** (Suppl. 3), 254 (1990).

84. E. J. Topol in E. J. Topol, ed., *Textbook of Interventional Cardiology*, W. B. Saunders Co., Philadelphia, Pa., 1991, p. 76.

85. R. A. G. Smith and co-workers, *Nature* **290**, 505 (1981).

86. H. Ferres, *Drugs* **33** (Suppl. 3), 33 (1987).

87. P. Doenecke and co-workers, *Klin. Wschr.* **64**, 682 (1986).

88. M. Kohler and co-workers, *Haemostasis* **17**, 32 (1987).

89. W. S. Hillis and co-workers, *Br. Heart J.* **53**, 178 (1985).

90. W. S. Hillis and co-workers, *Drugs* **33** (Suppl. 3), 117 (1987).

91. A. D. Timmis and co-workers, *JACC* **10**, 205 (1987).

92. S. Ikram and co-workers, *Br. Med. J.* **293**, 786 (1986).

93. M. Been and co-workers, *Int. J. Cardiol.* **11**, 53 (1986).

94. M. Been and co-workers, *Br. Heart J.* **53**, 253 (1985).

95. J. L. Anderson and co-workers, *J. Am. Coll. Cardiol.* **11**, 1153 (1988).

96. H. J. R. M. Bonnier and co-workers, *Am. J. Cardiol.* **62**, 25 (1988).

97. J. Lopez-Sendon and co-workers, *Circulation* **33** (Suppl. 2), 277 (abstr.) (1988).

98. M. L. Brouchier and co-workers, *Drugs* **33** (Suppl. 3), 140 (1987).

99. K. J. Hogg and co-workers, *Lancet* **335**, 254 (1990).

100. J. L. Anderson and co-workers, *Circulation* **83** (Suppl.), 152A (1991).

101. AIMS Trial Study Group, *Lancet* **i**, 545 (1988).

102. D. Collen, *J. Cell. Biochem.* **33**, 77 (1987).

103. W. E. Holmes and co-workers, *Bio/Technology* **3**, 923 (1985).

104. A. Riccio and co-workers, *Nucleic Acids Res.* **13**, 2759 (1985).

105. D. Collen, *Haemostasis* **43**, 77 (1980).

106. L. Neuhaus, *Cardiovasc. Intervent. Radiol.* **9**, 253 (1986).

107. S. N. Tennant and co-workers, *Circulation* **69**, 756 (1984).

108. M. Yasuno and co-workers, *Am. J. Cardiol.* **53**, 1217 (1984).

109. D. Cemigliaro and co-workers, *G. Ital. Cardiol.* **12**, 365 (1982).

110. K. Kammatsue and N. Kajwara, *Jpn. Circ. J.* **48**, 713 (1984).

111. H. Tateishi and co-workers, *Jpn. Circ. J.* **46**, 756 (1982).

112. C. Cemigliaro and co-workers, *G. Ital. Cardiol.* **14**, 927 (1984).

113. D. G. Mathey and co-workers, *Am. J. Cardiol.* **55**, 878 (1985).

114. T. Kohno and co-workers, *Bio/Technology* **2**, 628 (1984).

115. V. Gurewich and co-workers, *J. Clin. Invest.* **73**, 1731 (1984).

116. V. Gurewich and R. Pannell, *Thromb. Res.* **44**, 217 (1986).

117. R. Pannell and V. Gurewich, *Blood* **67**, 1213 (1986).

118. S. S. Hussain, B. Lipinski, and V. Gurewich, *Arch. Biochem. Biophys.* **220**, 31 (1983).

119. V. Gurewich, *J. Am. Coll. Cardiol.* **10**, 16B (1987).

120. F. Van de Werf, M. Nabuhara, and D. Collen, *Ann. Intern. Med.* **104**, 345 (1986).

121. W. Kasper and co-workers, *Klin. Wochenschr.* **66** (Suppl. 12), 109 (1988).

122. C. Bode and co-workers, *Am. J. Cardiol.* **61**, 971 (1988).

123. D. C. Gulba and co-workers, *Am. J. Cardiol.* **63**, 1025 (1989).

124. J. Loscalzo and co-workers, *Circulation* **79**, 776 (1989).

125. T. Astrup, *Fed. Proc. Am. Soc. Exp. Biol.* **42**, 25 (1966).

126. D. C. Rijken and co-workers, *Biochem. Biophys. Acta.* **140**, 580 (1979).

127. D. C. Rijken and D. Collen, *J. Biol. Chem.* **256**, 7035 (1981).

128. G. A. Vehar and co-workers, *Bio/Technology* **2**, 1051 (1984).

129. D. Pennica and co-workers, *Nature* **214**, 301 (1983).

130. G. Pohl and co-workers, *Biochemistry* **23**, 3701 (1984).

131. M. Hoylaerts and co-workers, *J. Biol. Chem.* **257**, 2912 (1982).

132. M. Ranby, *Biochem. Biophys. Acta.* **704**, 461 (1982).

133. A. J. Tiefenbrunn, A. K. Robison, and P. B. Kurnik, *Circulation* **71**, 110 (1985).

134. F. Van der Werf and co-workers, *N. Engl. J. Med.* **310**, 609 (1984).

135. D. Collen and co-workers, *J. Pharmacol. Exp. Ther.* **231**, 146 (1984).

136. European Cooperative Study Group, *Lancet* **i**, 842 (1985).

137. The TIMI Study Group, *N. Engl. J. Med.* **312**, 932 (1985).

138. E. J. Topol and co-workers, *J. Am. Coll. Cardiol.* **9**, 1205 (1987).

139. H. Mueller and co-workers, *J. Am. Coll. Cardiol.* **10**, 479 (1987).

140. A. D. Guerci and co-workers, *N. Engl. J. Med.* **317**, 1613 (1987).

141. M. O'Rourke and co-workers, *Circulation* **77**, 1311 (1988).

142. R. G. Wilcox and co-workers (ASSET Study Group), *Lancet* **ii**, 525, (1988).

143. F. Van der Werf and co-workers, *Br. Med. J.* **297**, 1374 (1988).

144. G. R. McKendall and co-workers, *J. Am. Coll. Cardiol.* **18**, 1774 (1991).

145. Gruppo Italiano pre lo Studio della Sopravvivenza nell'Infarto Miocardio (GISSI-2), *Lancet* **336**, 65 (1990).

146. The International Study Group, *Lancet* **336**, 71 (1990).

147. Third International Study of Infarct Survival Collaborative Group (ISIS-3 trial), *Lancet* **339**, 753 (1992).

148. P. Ott and P. Fenster, *Am. Heart J.* **121**, 1583 (1991).

149. V. Gurech and R. Pannell, *Thromb. Res.* **44**, 217 (1986).

150. R. Pannell and V. Gurech, *Blood* **67**, 1215 (1986).

151. D. Collen and co-workers, *Circulation* **74**, 838 (1986).

152. A. A. Ziskind and co-workers, *Clin. Res.* **35**, 337A (1987).

153. A. A. Ziskind and co-workers, *Circulation* **79**, 393 (1989).

154. C. Bode and co-workers, *Circulation* **81**, 907 (1990).

155. J. M. Kirshenbaum and co-workers, *Am. J. Cardiol.* **68**, 1564 (1991).

156. D. Collen, *Am. J. Cardiol.* **69**, 71A (1992).

157. C. L. Grines and co-workers, *J. Am. Coll. Cardiol.* **14**, 573 (1989).

158. S. L. Voytik and co-workers, *Thromb. Haemostasis* **64**, 438 (1991).

159. B. R. H. Farrow, *Cornell Vet.* **79**, 211 (1989).

160. K. Sandhoff in J. A. Barranger and R. O. Brady, eds., *Molecular Basis of Lysosomal Storage Disorders*, Academic Press, Inc., Orlando, Fla., 1984, p. 19.

161. D. E. George and J. A. Mangos, *J. Pediatr. Gastroenterol. Nutr.* **7**, S49 (1988).

162. V. L. Soutter and co-workers, *Clin. Gastroenterol.* **15**, 137 (1986).

163. A. M. Weber, C. C. Roy, and L. Chartrand, *Gut* **17**, 295 (1976).

164. K. Cox and co-workers, *J. Pediatr.* **94**, 488 (1979).

165. P. A. Durie and co-workers, *Gut* **21**, 778 (1980).

166. D. Y. Graham, *Dig. Dis. Sci.* **27**, 485 (1982).

167. A. P. V. Varro, *Hepato-gastroenterol.* **35**, 83 (1988).

168. L. I. Geller and M. M. Pashto, *Ter. Arkh.* **61**, 60 (1989).

169. I. M. Roberts, *Pancreas* **4**, 496 (1989).

170. D. W. Beverley and co-workers, *Arch. Dis. Child* **62**, 564 (1987).

171. E. G. Nassiff and co-workers, *J. Pediatr.* **98**, 320 (1981).

172. R. J. Stead and co-workers, *Thorax* **42**, 533 (1987).

173. M. S. Hershfield and co-workers, *N. Engl. J. Med.* **316**, 589 (1987).

174. Y. Levy and co-workers, *J. Pediatr.* **113**, 312 (1988).

175. J. A. Biller and co-workers, *J. Pediatr.* **111**, 91 (1987).

176. M. Moskovitz and co-workers, *Am. J. Gastroenterol.* **82**, 632 (1987).

177. Ref. 61, p. 90.

178. Ref. 61, p. 131f.

179. M. Kimura and co-workers, *Proc. Soc. Exp. Biol. Med.* **188**, 364 (1988).

180. R. D. Smyth, R. M. Brennan, and G. J. Martin, *Am. J. Pharm.* **133**, 294 (1961).

181. G. I. Zatuchni and D. J. Colombi, *Obstet. Gynecol.* **29**, 275 (1967).

182. A. Capria and M. Marchioro, *Minerva Gastroenterol.* **17**, 84 (1971).

183. N. W. Ahle, *Ann. Emerg. Med.* **16**, 1063 (1987).

184. N. Vespa, *Minerva Med.* **63**, 3219 (1972).

185. L. Smith, *J. Am. Med. Assoc.* **187**, 137 (1964).

186. J. A. Thomas and co-workers, *Clin. Orthrop.* **206**, 61 (1986).

187. J. R. Jenner, D. J. Buttle, and A. K. Dixon, *Ann. Rheum. Dis.* **45**, 441 (1986).

188. C. Watts and E. Dickhaus, *Surg. Neurol.* **26**, 236 (1986).

189. C. B. Shields and co-workers, *J. Neurosurg.* **67**, 187 (1987).

190. L. Smith and J. E. Brown, *J. Bone Joint Surg.* **49B**, 502 (1987).

191. B. S. Sussman, *J. Neurosurg.* **42**, 389 (1975).

192. C. Watts, R. Knighton, and G. Roulhac, *J. Neurosurg.* **42**, 374 (1975).

193. B. S. Sussman, J. W. Bromley, and J. C. Gomez, *J. Am. Med. Assoc.* **245**, 730 (1981).

194. M. D. Brown and J. S. Tompkins, *Spine* **11**, 123 (1986).

195. P. J. Pollack, *Corr. Ther. Res.* **4**, 229 (1962).

196. G. D. Magnes, *J. Am. Dent. Assoc.* **72**, 1920 (1966).

197. I. Mandl, *Azneim. Forsch/Drug Res.* **32**, 1381 (1982).

198. L. Yinqiu and co-workers, *J. Trauma* **28**, 5227 (1988).

199. S. Sherry and A. P. Fletcher, *Clin. Pharmacol. Ther.* **1**, 202 (1960).

200. E. Forsling, *Eur. J. Pharmacol.* **33**, 637 (1988).

201. K. C. Norris and co-workers, *Am. J. Kidney Dis.* **10**, 62 (1987).

202. H. M. Katzin, *Arch. Ophthalmol.* **35**, 415 (1945).

203. N. A. Puchkovskay and T. U. Gorgiladze, *Cornea* **6**, 70 (1987).

204. P. M. McCarthy and co-workers, *Ann. Thorac. Surg.* **45**, 203 (1988).

205. S. Lippert and S. Borgeskov, *Ugeskr. Laeger* **149**, 3328 (1987).

206. H. W. Wacławiczek, F. Chmelizek, and I. Koller, *Surg. Endosc.* **1**, 99 (1987).

207. K. Baumgarten and S. Moser, *J. Perinat. Med.* **14**, 43 (1986).

208. Y. M. Goudarzi, *Akt. Traumatol.* **16**, 186 (1986).

209. H. Groitl and J. Scheele, *Surg. Endosc.* **1**, 93 (1987).

210. J. H. Lemmer and co-workers, *J. Pediatr. Surg.* **22**, 829 (1987).

211. M. Abbildungen, *Zent. Bl. Chir.* **112**, 99 (1987).

212. A. Narakas, *Orthop. Clin. North Am.* **19**, 187 (1988).

213. R. Pelligrini and P. Vartova, *Azneim. Forsch.* **19**, 149 (1975).

214. W. F. Rathgeber, *S. Afr. Med. J.* **45**, 185 (1971).

215. G. Gaspardy and co-workers, *Rheumatol. Phys. Med.* **11**, 14 (1971).

216. P. C. Shaw, *Br. J. Clin. Pract.* **23**, 25 (1969).

217. J. W. Finch, *J. Okla. State Med. Assoc.* **79**, 766 (1986).

218. V. I. Pronin and co-workers, *Meditsin. Radiol (Mosk.)* **33**, 29 (1988).

219. J. M. Oliverra, R. Carballo, and H. A. Zimmerman, *Am. Heart J.* **57**, 712 (1959).

220. I. G. Calder and V. H. Smith, *Brit. J. Ophthalmol.* **70**, 418 (1986).

221. S. R. Hein and R. H. Keates, *Ophthalmic Surg.* **17**, 731 (1986).

222. S. N. Bhagwati and K. George, *Child's Nerv. Syst.* **2**, 20 (1986).

223. B. J. Few, *Am. J. Maternal Child Nursing* **12**, 23 (1987).
224. L. Frankenberg, *Arch. Toxicol.* **45**, 315 (1980).
225. S. Carson and co-workers, *Toxicol. Appl. Pharmacol.* **26**, 184 (1977).
226. H. Witschi, *Fed. Proc.* **35**, 1631 (1977).
227. N. A. Rao and co-workers, *Ophthalmic Res.* **18**, 41 (1986).
228. G. E. Marek and co-workers, *Ophthalmic Res.* **17**, 297 (1985).
229. N. A. Rao and co-workers, *Invest. Ophthal. vis. Sci.* **28**, 886 (1987).
230. L. Flohüi, *Mol. Cell. Biochem.* **84**, 123 (1988).
231. A. Petkau, *Mol. Cell. Biochem.* **84**, 133 (1988).
232. F. Baillet and co-workers, *Free Radic. Res. Commun.* **1**, 387 (1986).
233. K. H. Lim and co-workers, *Ann. Thorac. Surg.* **42**, 282 (1986).
234. N. A. Rao and co-workers, *Ophthalmic Res.* **20**, 213 (1988).
235. P. H. Sato and I. V. Grahm, *Arch. Biochem. Biophys.* **210**, 609 (1981).
236. M. Miura, *Jpn. J. Nephrol.* **29**, 1189 (1987).
237. M. Miura and co-workers, *Clin. Nephrol.* **32**, 209 (1989).
238. T. M. S. Chang, *Methods Enzymol.* **137**, 444 (1988).
239. Y. Shimizu and co-workers, *Biomater. Med. Devices Artif. Organs* **5**, 49 (1977).
240. Y. Shimizu and co-workers, *Biomater. Med. Devices Artif. Organs* **6**, 375 (1988).
241. Y. Shimizu and co-workers in G. D. Winter, D. F. Gibbons, and H. Plenck, Jr., eds., *Biomaterials 1980*, John Wiley & Sons, Inc., New York, 1982, p. 745.
242. U. Sprandel, *Bibithca Haemat.* **51**, 7 (1985).
243. M. J. Poznansky, *Methods Enzymol.* **137**, 566 (1980).
244. M. J. Poznansky and D. Bharadwaj, *Can. J. Physiol. Pharmacol.* **58**, 322 (1980).
245. M. J. Poznansky and co-workers, *Science* **223**, 1304 (1984).
246. E. A. Wolfe and T. M. S. Chang, *Int. J. Artif. Organs* **10**, 269 (1987).
247. R. Pool, *Science* **248**, 305 (1990).
248. H. Hyden, *Biomater. Med. Devices Artif. Organs* **8**, 1 (1980).
249. R. J. Desnick, *Adv. Exp. Med. Biol.* **101**, 753 (1978).
250. M. B. Fiddler and co-workers, *J. Lab. Clin. Med.* **96**, 307 (1980).
251. S. R. Thorpe and co-workers, *Pediat. Res.* **9**, 918 (1975).
252. "Drug Topics," *The Red Book*, Medical Economic Data, Montavale, N.J., 1992.
253. R. L. Capizzi and Y-C. Chang in J. S. Holcenberg and J. Roberts, eds., *Enzymes as Drugs*, John Wiley & Sons, Inc., New York, 1981, pp. 1-24.
254. J. Ring and K. Messmer, *Lancet* **1**, 466 (1977).
255. *United States Pharmacopeia XIX*, Mack Publishing Co., Easton, Pa., 1975.

Hiremagalur N. Jayaram
Indiana University School of Medicine

Gurpreet S. Ahluwalia
Gillette Research Institute

David A. Cooney
National Cancer Institute

ENZYME DETERGENTS.

See DETERGENCY; ENZYME APPLICATIONS, INDUSTRIAL.

ENZYME INHIBITORS

The development of potent enzyme inhibitors has led to an increased understanding of enzyme mechanisms and has provided effective therapeutic agents for the treatment of diseases. Enzymes are natural biocatalysts that promote specific reactions essential for the viability of living organisms. They have unique recognition sites that allow them to select their substrates out of the vast pool of biologically important compounds in living cells. The substrate binds to the active site, that portion of the enzyme responsible for promoting the chemistry involved in converting the substrate to product. Inhibitors of enzymes prevent this chemical reaction from occurring by altering the active site and thereby rendering the enzyme at least temporarily inactive. Generally, the more tightly the inhibitor binds to the active site, the better; but it also should be relatively specific for the target enzyme. An increased understanding of the enzyme's specificity for substrate and inhibitor binding enables a more rational design of potent inhibitors, selective for a particular enzyme. There has been movement away from the development of enzyme inhibitors by screening of natural products to the more efficient approach of so termed rational drug discovery. This article presents several different strategies for rational drug discovery based on enzyme inhibition. It focuses on the basic concepts and classifies enzyme inhibitors. Other related in-depth reviews have appeared (1–4).

Applications for Enzyme Inhibitors

Enzyme inhibitors are often used to further the understanding of enzyme mechanisms. Inhibitors serve as probes for kinetic and chemical processes during catalysis. For example, they help elucidate the mode of ligand binding, where a ligand is a compound like an inhibitor, substrate, or substrate analogue that can bind to the active site of an enzyme. Alternative applications for inhibitors are the detection of short-lived enzyme-bound reaction intermediates, or the identification of amino acid residues at the active site that are necessary for the catalytic activity of the enzyme. Inhibitors are also used for *in vivo* studies to localize and quantify enzymes in organs or to mimic certain genetic diseases that involve the absence of an enzyme in a given biosynthetic pathway.

In pharmacological research, enzyme inhibitors are used to inactive specific enzymes or groups of enzymes, leading to the treatment of many diseases. In order for an enzyme inhibitor to be an effective drug, it not only has to be a potent inhibitor for its specific target enzyme but must also have good bioavailability, enabling the inhibitor to reach its target of inhibition in effective therapeutic concentrations. Therefore, drugs should have a high specificity toward the target enzyme in order to avoid the depletion of the inhibitor concentration in the host by nonspecific pathways. If a drug only interacts with the target enzyme and not with other sites in the body, side-effects caused by inhibition of vital enzymes and formation of toxic decomposition products will be minimal. A useful drug must also be delivered to its site of action in the body. For example, compounds that are too polar and highly charged are usually unable to cross cell membranes readily, and are therefore generally not useful for the inhibition of cellular enzymes *in vivo*.

An important area of drug design is the development of drugs against microorganisms and parasites in humans. A powerful strategy consists of choosing a target enzyme that is essential for the existence of the invader, but not the host. Therefore, these drugs usually exhibit low toxicity toward the host. A classic example of a target enzyme for the design of antibacterial agents (qv) is alanine racemase, an enzyme that does not exist in humans. An alternative approach is the choice of a target enzyme that exists as different species, called isoenzymes, in the invader and the host. Even though the two isoenzymes are highly similar, they usually have at least some subtle structural differences. Selective inhibitors can then be developed that bind with much higher affinity to the invader's version of the enzyme.

Another class of therapeutic agents is used for the treatment of certain genetic diseases or other enzymatic disorders caused by the dysfunction or absence of one particular enzyme. This often leads to an unwanted accumulation or imbalance of metabolites in the organism. For example, some anticonvulsive agents are inhibitors for γ -aminobutyric acid aminotransferase [9037-67-6]. An imbalance of two neurotransmitters, glutamate and γ -aminobutyric acid, is responsible for the symptoms. Inhibition of the enzyme leads to an increase of its substrate γ -aminobutyric acid, decreasing the imbalance and subsequently relieving the symptoms of the disease.

An important drug in the regulation of cholesterol metabolism is lovastatin [75330-75-5] which is an HMG–CoA reductase inhibitor (see CARDIOVASCULAR AGENTS). β -Hydroxy- β -methyl glutarate–coenzyme A (HMG–CoA) reductase is the rate-limiting enzyme of cholesterol synthesis. Lovastatin is actually a prodrug, which is eventually hydrolyzed in the liver to its active, β -hydroxylated form (5).

A third class of enzyme inhibitors consists of antitumor agents. Their design is quite challenging because tumor cells do not to appear to contain enzymes that are very different from those in normal cells. Most antitumor drugs, called antiproliferative agents, take advantage of the fact that tumor cells generally grow and divide much faster than normal cells. However, since normal cells also still continue to divide, this class of agents exhibits significant toxicity in the host. In particular, the bone marrow stem cells and epithelial cells are affected because of their relatively high rates of growth (see CHEMOTHERAPEUTICS, ANTICANCER).

Classification of Enzyme Inhibitors

The enzyme inhibitors discussed in this article may be classified as follows:

Noncovalent inhibitors	Covalent inhibitors
rapid reversible inhibitors	mechanism-based inhibitors
tight, slow, slow-tight binding inhibitors	affinity labels
transition-state analogues	pseudoirreversible inhibitors
multisubstrate analogues	

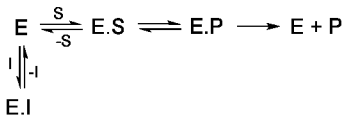
All of these enzyme inhibitors are active site-directed. They bind to the substrate binding site where the catalysis specific for that particular enzyme takes place. Inhibitors binding to a site other than the active site of the enzyme are called allosteric effectors, allosteric inhibitors, or allosteric cofactors, and are not considered here. The categories of noncovalent inhibitors may appear arbitrary because certain inhibitors may actually fit into more than one of the subcategories. This situation arises because the terminology for each class describes different properties of the inhibitors such as kinetic behavior, mechanism of action, or structure. Therefore, the subcategories cannot be compared with one another, and are only used to illustrate the rationale behind each design strategy. A number of good reviews on the general design of enzyme inhibitors have appeared (2).

NONCOVALENTLY BINDING ENZYME INHIBITORS

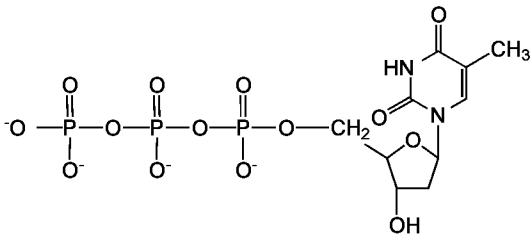
These inhibitors bind to the enzyme without forming a covalent bond. Their affinity for the active site of the enzyme is determined by the degree of electrostatic and hydrophobic attraction forces and hydrogen bonds between the enzyme and the ligand. The subcategories for this class of inhibitors are founded on different criteria. Based on their kinetic behavior, one can distinguish between classical rapidly reversible, slow-binding, tight-binding, and slow, tight-binding inhibitors. Other classes of noncovalent inhibitors are multisubstrate and transition-state analogues, which mimic the interactions of reaction intermediates or transition-states with the active site of the enzyme.

Rapidly Reversible Inhibitors. Some inhibitors can be described solely by their kinetics. Rapidly reversible inhibitors, which are classified by their Michaelis-Menten kinetics (1,6,7) specifically, are so named because they rapidly and reversibly establish their binding equilibria with enzymes. They bind to enzymes noncovalently, obtaining binding energy from thermodynamic sources such as the hydrophobic effect, van der Waals contacts (induced dipole interactions), hydrogen-bond formation, and charge–charge interactions (1,6–8). The strength of enzyme-inhibitor binding is measured by the K_i value, defined as the inhibition constant or dissociation constant of the inhibitor from the enzyme-inhibitor complex. A low K_i value represents a tightly bound enzyme-inhibitor complex. Another measurement of inhibition is the IC_{50} , an approximation of the K_i , which is the inhibitor concentration that produces 50% enzyme inhibition. More detailed discussion of K_i values can be found in general texts (1,6,7). The rapidly reversible class of inhibitors is comprised of competitive, noncompetitive, and uncompetitive inhibitors. These types of inhibitors are distinguished by conducting competition experiments, which determine how an inhibitor interacts with the substrate at the active site of the enzyme.

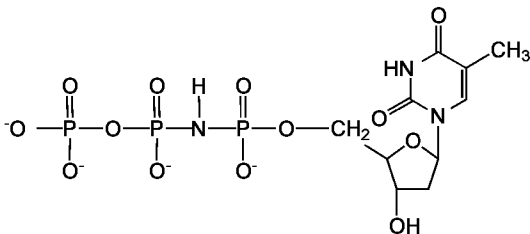
A competitive inhibitor binds to the same enzyme form as the substrate, competing for the same site, such that the inhibitor and substrate prevent one another from binding. This may be represented as follows:*Competitive inhibition*



A recent example is the substrate analogue thymidine 5'-[α,β-imido]triphosphate [141171-20-2] (TMPNPP) (2) which competitively inhibits the human immunodeficiency virus-1 (HIV-1) reverse transcriptase (HIV-1 RT) with a K_i value of 2.4 micromolar (μM) (9). The substrate is thymidine 5'-triphosphate TPP [365-08-2] (1).

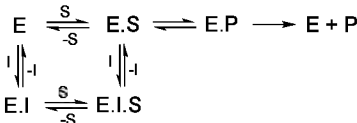


(1)



(2)

Unlike a competitive inhibitor, a noncompetitive inhibitor does not compete with the substrate for the binding site, since the inhibitor and substrate can bind to the enzyme either independently or simultaneously.*Noncompetitive inhibition*



It has been shown for porcine pancreatic α-amylase how the kinetic classification of an inhibitor can be influenced by the size of the substrate (10). Substrate and inhibitor binding sites in the active site of the enzyme were proposed based on a series of competition experiments (10). With a small substrate, G₂PNP (3), the inhibitor 4-phenylimidazole [670-95-1] (5) shows noncompetitive kinetics with a K_i value of 7.5 mM (Fig. 1). However, with a large substrate, G OH (4), 4-phenylimidazole shows competitive kinetics ($K_i = 7.1$ mM). When the substrate is small, the inhibitor binds simultaneously with the substrate at the active site; however, when the substrate is large, the inhibitor can no longer bind simultaneously because the larger substrate occupies the inhibitor's binding site. With noncompetitive inhibition, the substrate and inhibitor bind to the enzyme with equal strength; however, with mixed noncompetitive inhibition the binding strengths, or affinities, are different.

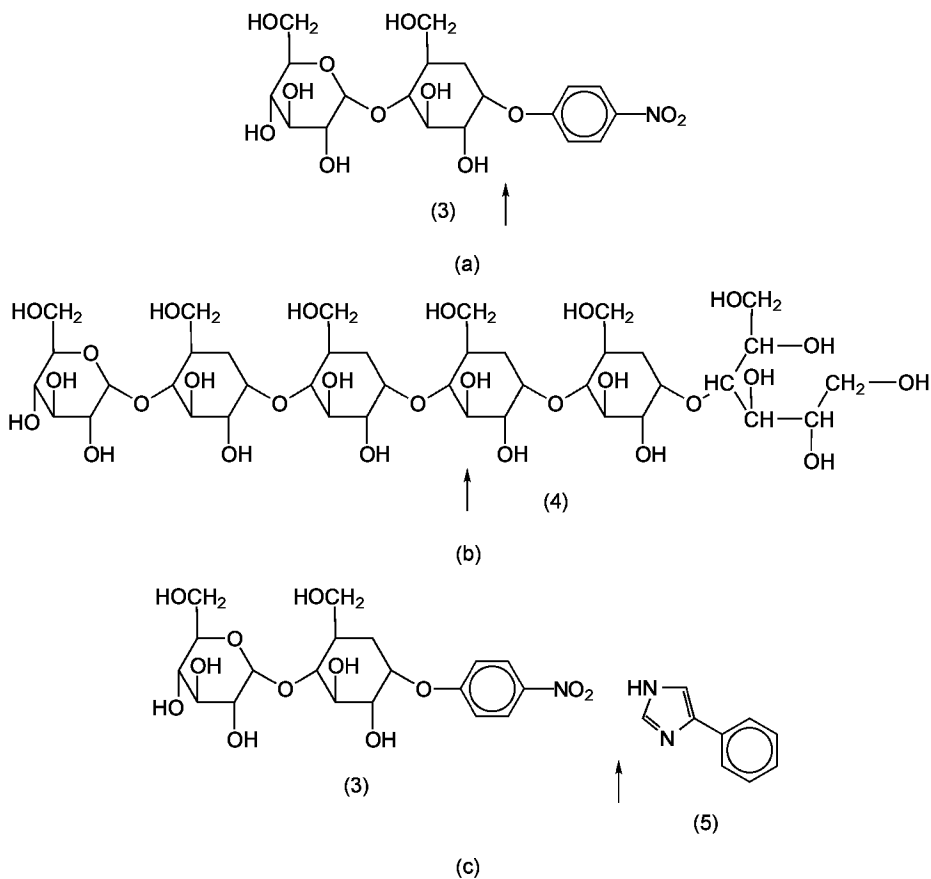
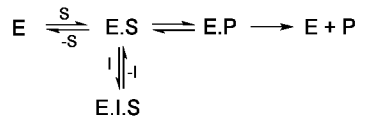
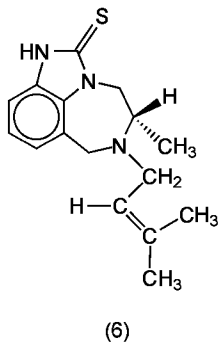


Fig. 1. Inhibition of porcine pancreatic α -amylase. Substrates, an inhibitor, and their binding orientations in the active site are shown schematically. The arrows denote the catalytic site in each case. (a) The small substrate, G_2 PNP [17400-77-0] (3); (b) the large substrate, G OH [13532-61-1] (4); and (c) the inhibitor, 4-phenylimidazole (5) and the substrate G_2 PNP (3) in the binding orientation for noncompetitive inhibition. The binding orientation of G_2 PNP is different than it is in the absence of inhibitor.

Like a noncompetitive inhibitor, an uncompetitive inhibitor does not compete with the substrate since it binds to the enzyme–substrate complex but not to the free enzyme. *Uncompetitive inhibition*



An example has been published (11) in which TIBO R82150 [126320-77-2] (6), an inhibitor of the HIV-1 RT, is noncompetitive with respect to one substrate, deoxyguanosine triphosphate ($dGTP$), yet uncompetitive with respect to the second, the template–primer complex $\text{poly(C)} \cdot (dG)_{12-18}$. The IC_{50} of (6) is $0.34 \mu M$.

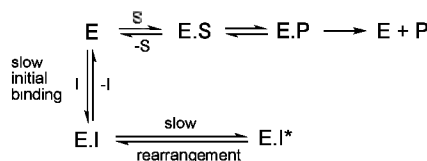


Tight, Slow, and Slow, Tight-Binding Inhibitors. These inhibitors can be distinguished from the classical, rapidly reversible inhibitors because they do not display Michaelis-Menten kinetics. Each of the different inhibitors can be identified by its concentration relative to that of the enzyme and the rate of equilibrium formation during the reaction (Table 1). Tight-binding inhibitors inhibit reactions at concentrations comparable to that of the enzyme and under conditions where the equilibrium of complex formation sets up rapidly. Tight-binding, noncovalent inhibitors may have such high affinities for their target enzymes that they become functionally equivalent to covalent or irreversible inhibitors. These compounds may have half-lives on the order of hours, days, or even months. They have very low K_i values (eg, picomolar [pM]), and their off-rates are very slow. In the cases where one equivalent of inhibitor binds irreversibly to one equivalent of enzyme, the inhibitors have been called stoichiometric. Such inhibitors usually show noncompetitive kinetics even though they compete with the substrate for the active site (12).

Table 1. Reversible Enzyme Inhibitors

Inhibitor class	I : E ratio	Attainment of equilibrium $E + I \rightleftharpoons EI$
classical	$I \gg E$	fast
tight binding	$I \approx E$	fast
slow binding	$I \gg E$	slow
slow, tight-binding	$I \approx E$	slow

Slow-binding inhibitors operate by one of two mechanisms. Either the inhibitor binds slowly in an initial step, or the initial binding step occurs quickly, followed by a slow rearrangement of the E·I complex.

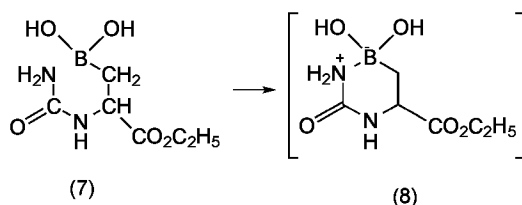


Most examples in the literature follow the slow rearrangement route. Slow binding can result when the form of the inhibitor or enzyme required for binding is present in low concentrations in solution, or if barriers to binding are encountered at the binding site (slow initial binding). Slow binding has also been correlated to the extrusion of an enzyme-bound water molecule from the active site after the formation of an initial enzyme-inhibitor complex (slow rearrangement) (13). Slow binding may arise from conformational changes in the enzyme-inhibitor complex for the formation of the activated complex at the transition-state (slow rearrangement) (12).

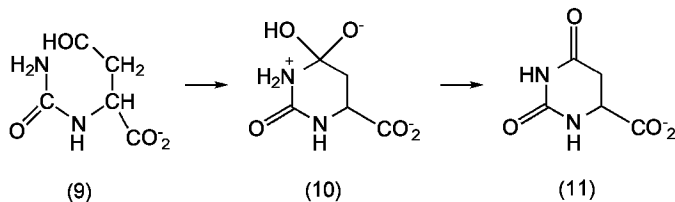
Slow, tight-binding inhibition occurs when slow-binding inhibition takes place at inhibitor concentrations comparable to that of the enzyme, in which case the previous two mechanisms can still apply. Comprehensive review articles on the subject of tight, slow, and slow, tight-binding inhibitors are available in the literature (12,14).

Transition-State Analogues. These are stable compounds that resemble the structures of the activated complexes existing at the transition-states of enzymatic reactions. These transition-states represent energy maxima during the catalysis, and the activated complexes are the short-lived, unstable structures at these energy maxima. Basic knowledge of the enzymatic reaction mechanism is necessary in order to design a transition-state analogue. The potency of these analogues results from the extra favorable binding interactions that appear in the transition-state but not in the ground-state complex. These interactions lead to much tighter binding by lowering the energy of activation required for the reaction. Furthermore, these inhibitors bind more rapidly than substrates because they avoid processes like bond rehybridization, and they reduce molecular entropy. However, most of the designed transition-state analogues may more closely resemble an actual reaction intermediate rather than an activated complex at the transition-state because activated complexes have partial bonds and are highly charged, whereas a transition-state analogue must be stable and have fully developed bonds. However, both reaction intermediates and transition-state analogues take advantage of the higher binding affinity of the enzyme for these high-energy species (15–20).

Structure 8 (a charged form of (7) [124419-33-6]) represents a well designed transition-state analogue for dihydroorotase (21).



The mechanism of catalysis of dihydroorotase may be outlined as follows:



The reaction is proposed to proceed from the anion (9) of *N*-aminocarbonylaspartic acid [923-37-5] to dehydroorotate (11) via the tetrahedral activated complex (10), which is a highly charged, unstable *sp*³ carbon species. In order to design a stable transition-state analogue, the carboxylic acid in dihydroorotate (hexahydro-2,6-dioxo-4-pyrimidinecarboxylic acid) [6202-10-4] was substituted with boronic acid; the result is a competitive inhibitor of dihydroorotase with a *K_i* value of 5 μ*M*. Its inhibitory function is supposedly due to the formation of the charged, but stable, tetrahedral transition-state intermediate (8) at the active site of the enzyme.

Multisubstrate Inhibitors. The design of multisubstrate inhibitors is suitable for the inactivation of enzymes that require the simultaneous binding of two or more substrates during catalysis. In order for catalysis to proceed, substrates must usually have particular orientations in the active site as well as close proximity to each other. Multisubstrate inhibitors take advantage of these requirements by mimicking the simultaneously bound substrates at the active site. Inhibitors that combine two substrates are termed bisubstrate analogues, those that combine three substrates are termed trisubstrate analogues, etc. These inhibitors are often highly specific because the combination of two or more substrates or products makes their structures unique. It is much less likely that a multisubstrate inhibitor will be recognized by other enzymes, as frequently occurs for single substrate analogue inhibitors. Multisubstrate inhibitors also have an increased binding affinity for the target enzyme, because the free energy of binding should be roughly the product of the free energies of each independent substrate. For example, if the two substrates of an enzymatic reaction have binding constants in the millimolar (*mM*) range, a bisubstrate analogue should have an approximate *K_i* value in the μ*M* range. An additional amount of free binding energy is gained due to the entropic advantage of reduced molecularity (8,16,22–24).

A good example for a multisubstrate inhibitor for glycylamide ribonucleotide transformylase has been provided (25). The enzyme catalyzes the transfer of the formyl group from 6*R*-CHO-H₄folate [68538-85-2] (13) to the substituted ribose [10074-18-7] (12) (Fig. 2). The multisubstrate inhibitor β-thioglycinamide ribonucleotide dideazafolate [119637-95-5] (15) was designed by linking substrates (12) and (13) with a stable thioether linkage and making a slight modification in the folate portion. It turned out to be the most potent inhibitor of this enzyme found so far, having a *K_i* value of 250 p*M*.

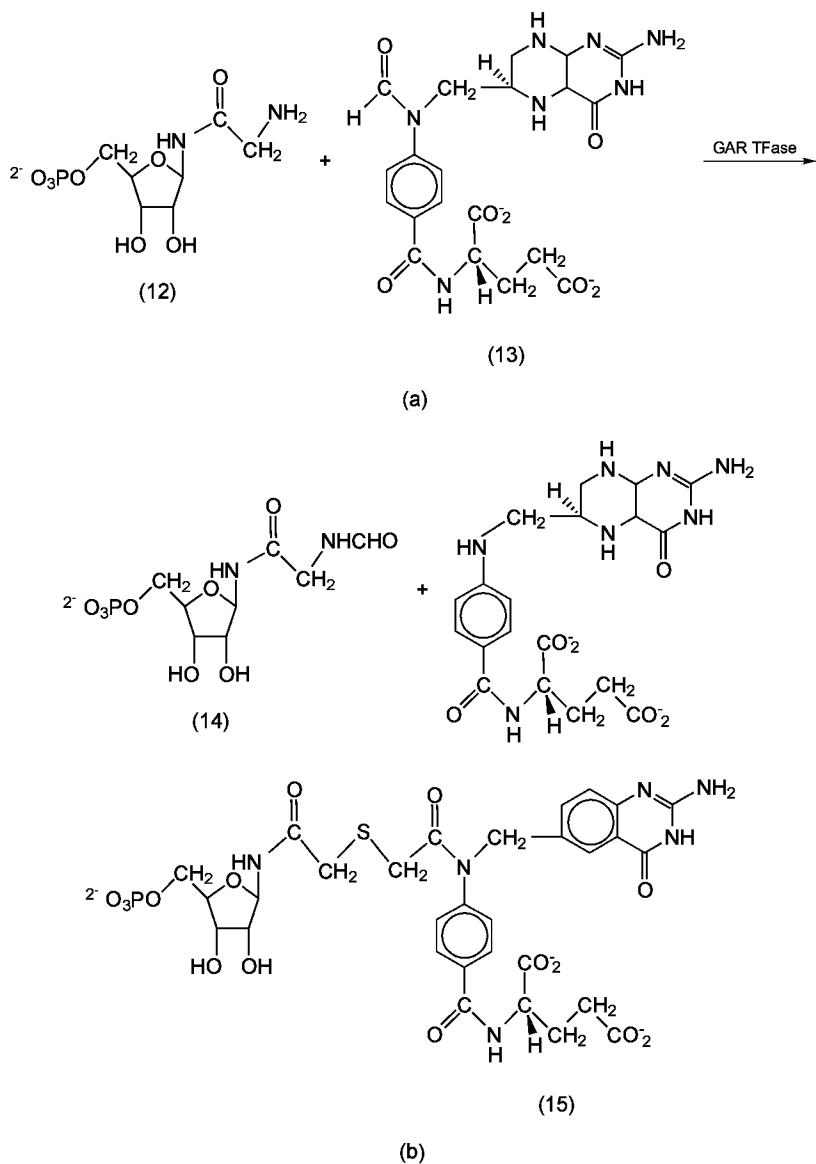


Fig. 2. Enzyme catalysis by glycinamide ribonucleotide transformylase (GAR TFase). **(a)** Transfer of the formyl group from (13) to (12) to form 2-(formylamino)-N-(5-O-phosphono-β-D-ribofuranosyl acetamide [349-34-8] (14); and **(b)** multisubstrate inhibitor (15).

COVALENTLY BINDING ENZYME INHIBITORS

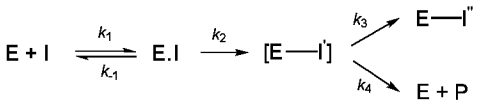
This class of inhibitors usually acts irreversibly by permanently blocking the active site of an enzyme upon covalent bond formation with an amino acid residue. Very tight-binding, noncovalent inhibitors often also act in an irreversible fashion with half-lives of the enzyme-inhibitor complex on the order of days or weeks. At these limits, distinction between covalent and noncovalent becomes functionally irrelevant. The mode of inactivation of this class of inhibitors can be divided into two phases; the inhibitors first bind to the enzyme in a noncovalent fashion, and then undergo subsequent covalent bond formation.

Mechanism-Based Inactivators. These inhibitors usually consist of an active site-directed, unreactive substrate or product analogue, containing a latent functional group that becomes activated during the normal catalysis of the enzyme. A reactive species is formed that then reacts with one of the enzyme's active site amino acid residues to form a covalent bond between the enzyme and the inhibitor. Commonly susceptible amino acid residues originate from cysteine, histidine, serine, threonine, lysine, aspartic acid, glutamic acid, and tyrosine. The design of mechanism-based inhibitors requires some knowledge of the catalytic mechanism of the enzyme with its normal substrate, and the introduction of an appropriate latent functional group into the substrate. The designed inhibitor must also fulfill the binding specificity requirements for the ligand recognition site. Furthermore, for covalent binding, the inhibitor requires a susceptible amino acid residue or cofactor in the active site in close proximity to the formed reactive group.

Inhibitors from this category are usually very specific toward their target enzymes. The fairly unreactive species is converted into the highly reactive species only during the specific catalytic step at the active site of the target enzyme. Subsequently, such inhibitors should generally show little cell toxicity, making them potentially useful as drugs. However, toxicity can still occur if some of the reactive species fail to bind to the enzyme and are released from the active site into solution. The reactive species may then react with other enzymes or turn into a noncovalent inhibitor upon hydrolysis.

The efficiency of inactivation by covalent bond formation vs release of the reactive species into solution has been described by its partition ratio. The most efficient inactivators have catalytic partition ratios of 0, in which case each inhibitor molecule leads to inactivation of the enzyme. To this date, many of these inhibitors have been designed, and alternative names like suicide substrate, Trojan Horse inactivator, enzyme induced inactivator, k_{cat} inhibitor, and latent inactivator have been used for this class of inhibitors. A number of comprehensive reviews are available (26–32).

A general kinetic scheme for mechanism-based inactivators is outlined as follows:



Usually, a rapid binding step of the inhibitor I to the enzyme E leads to the formation of the initial noncovalent enzyme-inhibitor complex E·I. This is usually followed by a rate determining catalytic step, leading to the formation of a highly reactive species [E-I']. This species can either undergo reaction with an active site amino acid residue of the enzyme to form the covalent enzyme-inhibitor adduct E-I'', or be released into the medium to form product P and free active enzyme E.

An evaluation of potential mechanism-based inactivators requires the fulfillment of the following criteria: (1) irreversible, active site-directed

inactivation of the enzyme upon the formation of a stable covalent linkage with the activated form of the inhibitor, (2) time- and concentration-dependent inactivation showing saturation kinetics, (3) involvement of a catalytic step, (4) inactivation prior to release of the activated species in order to prevent binding to locations other than the active site of the enzyme, and (5) a binding stoichiometry of 1:1 of inhibitor to enzyme's active site. A good treatment of the theoretical and practical aspects of the evaluation of mechanism-based inhibitors has been presented (29).

The underlying principles of mechanism-based inactivators have been illustrated (33) with the mechanism-based inactivation of alanine racemase from *Escherichia coli* B (Fig. 3). This pyridoxal phosphate-dependent enzyme (16) normally acts by forming a Schiff's-base type linkage with the α-amino group of alanine [6898-94-8] (step 1), followed by proton abstraction to form a resonance stabilized carbanion (step 2). The catalysis proceeds by reprotonation from the opposite side of the sp²-hybridized Schiff-base, resulting in an inversion of the stereochemistry (step 3). D-Chlorovinylglycine [114115-57-0] was found to be a very efficient and potent inhibitor of alanine racemase. A simplified version of its proposed mechanism of inactivation is outlined in Figure 4. First, it undergoes formation of a Schiff-base type linkage (step 1), followed by proton abstraction leading to the formation of an allenic intermediate (step 2). This complex is a strong electrophile capable of reacting with an active site nucleophile to form a covalent bond between the enzyme and the inhibitor.

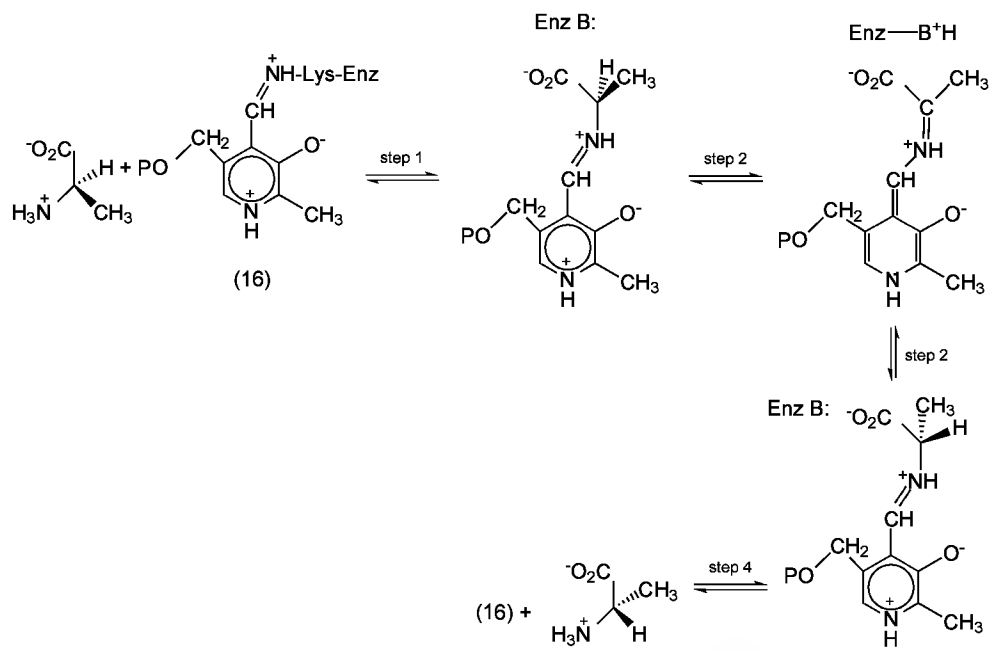


Fig. 3. Mechanism of catalysis of alanine racemase, P = PO²⁻.

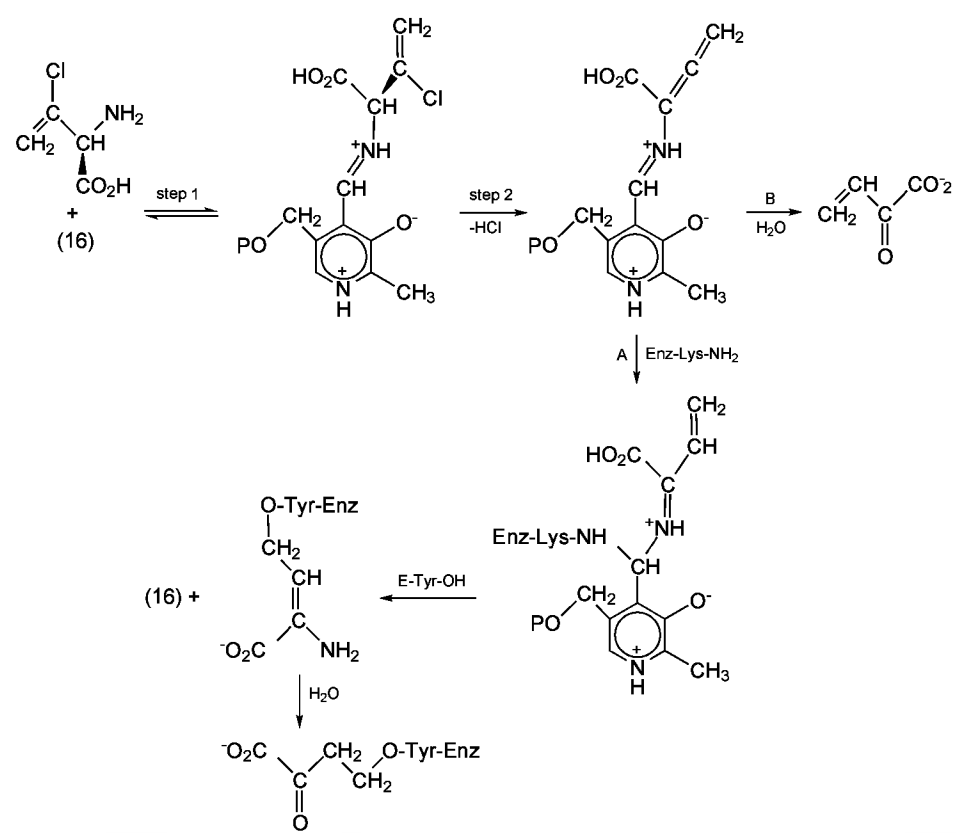


Fig. 4. Simplified mechanism-based irreversible inactivation of alanine racemase by 3-chlorovinylglycine, P = PO²⁻.

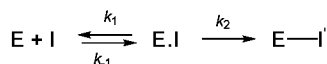
Affinity Labels. Active site-directed, irreversible inhibitors or affinity labels are usually substrate analogues that contain a reactive electrophilic functional group. In the first step, they bind to the active site of the target enzyme in a reversible fashion. Subsequently, an active site nucleophile in close proximity reacts with the electrophilic group on the substrate to form a covalent bond between the enzyme and the inhibitor, typically via S_N2 alkylation or acylation. Affinity labels do not require activation by the catalysis of the enzyme, as in the case of a mechanism-based inhibitor.

The design of potent inactivators requires the determination of the functional groups necessary for optimal binding of the inhibitor to the enzyme. Next, the regions of steric tolerance are identified that will be useful for introduction of the reactive group. The reactivity as well as the size of the reactive

group are critical for obtaining effective active site-directed irreversible inhibitors. Usually, by synthesizing and testing various substrate analogues, such optimal parameters are determined.

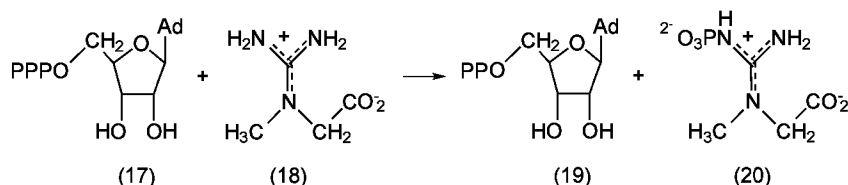
Affinity labels have the potential for good selectivity if they have an especially high affinity for their target enzyme. Once the inhibitor is bound to the active site, the unimolecular reaction of the bound inhibitor with an amino acid residue in close proximity to the reactive group at the active site is kinetically favored over the otherwise bimolecular reaction with other sites on the target enzyme or on other proteins. However, the principal pitfall with these inhibitors is the continuous presence of their reactive groups. They may not be useful *in vivo* because the target enzyme represents only a small fraction of the total number of molecules, and the reactive group may undergo bimolecular reactions with these other proteins, leading to toxicity.

The general kinetics for this class of inhibitors are outlined as follows:

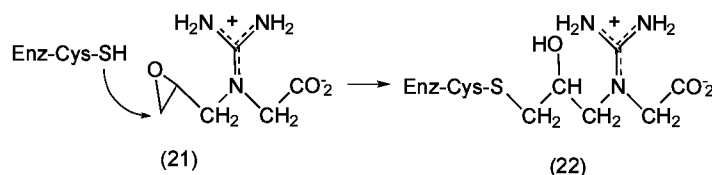


The often fast binding step of the inhibitor I to the enzyme E, forming the enzyme inhibitor complex E·I, is followed by a rate-determining inactivation step to form a covalent bond. The evaluation of affinity labels is based on the fulfillment of the following criteria: (1) irreversible, active site-directed inactivation of the enzyme upon the formation of a stable covalent linkage with the activated form of the inhibitor, (2) time- and concentration-dependent inactivation showing saturation kinetics, and (3) a binding stoichiometry of 1:1 of inhibitor to the enzyme's active site (34).

A good example of an affinity label for creatine kinase has been presented (35). This enzyme catalyzes the reversible transfer of a phosphoryl group from adenosine triphosphate [56-65-5] (17) to creatine [57-00-1] (18), leading to adenosine diphosphate [7384-99-8] (19) and phosphocreatine [67-07-2] (20).



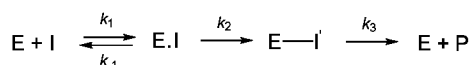
N-(2,3-Epoxypropyl)-*N*-amidinoglycine [70363-44-9] (21) was shown to be an affinity label of creatine kinase. Its mechanism of covalent bond formation is outlined as follows:



The attack of cysteine-282 at the less hindered methylene carbon was supported by model studies.

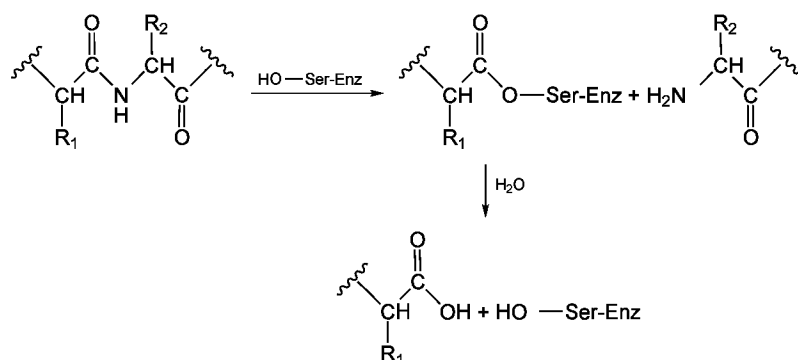
Pseudoirreversible Inactivators. These inhibitors behave as quasi-substrates since they are accepted as substrates upon binding to the active site of the enzyme. Subsequently, the enzyme starts its normal mechanism of catalysis, usually involving elimination-addition or acylation reactions. This leads to the creation of an electrophilic site capable of reacting with an active site amino acid to form a covalent bond between the enzyme and the inhibitor. Pseudoirreversible inhibitors neither contain reactive functionalities like affinity labels nor have latent functional groups that are converted into highly reactive species. The electrophilic species formed is not very reactive, and the adduct between the enzyme and the inhibitor is more like a reaction intermediate or steady state analogue, as observed for the normal substrates. Therefore, the covalent bond is not very stable, and may undergo either hydrolysis or simple reversal of the covalent bond formation, regenerating the active enzyme. The potency of this type of inhibitor and, therefore, their potential as drugs, depends on the half-life of the covalent adduct as well as the ease of the bond formation.

Criteria very similar to those for mechanism-based inactivators apply for the evaluation of pseudoirreversible inhibitors. The kinetics for pseudoirreversible inactivators are as follows:



A significant difference between pseudoirreversible inhibitors and mechanism-based inactivators is the reversibility of the inactivation. A complete evaluation of the mechanism involved would require evidence not only for the covalent enzyme-inhibitor complex, but also for its decomposition products and its rate of reactivation. It is often difficult to identify the active site amino acid residue covalently linked to the inhibitor because of the instability of the complex.

An example of a pseudoirreversible inhibitor has been demonstrated for chymotrypsin (36). This enzyme is a serine protease, and its mechanism of catalysis may be outlined as follows, where R₁ or R₂ preferentially is a hydrophobic amino acid residue.



The enzyme catalyzes the hydrolysis of an amide bond linkage with water via a covalent enzyme-inhibitor adduct. Benzoxazinones such as 2-ethoxy-4*H*-3,1-benzoxazin-4-one [41470-88-6] (23) have been shown to completely inactivate the enzyme in a competitive and stoichiometric fashion (Figure 5). The intermediate (25) is relatively stable compared to the enzyme-substrate adduct due to the electron-donating properties of the ortho substituents. The complex (25) has a half-life of reactivation of 11 hours.

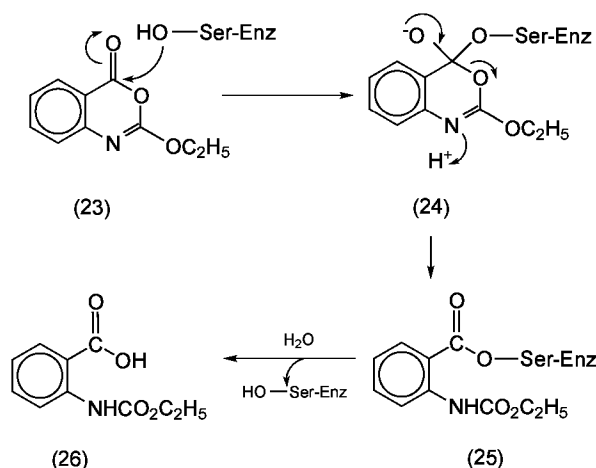


Fig. 5. Proposed mechanism of inactivation of α -chymotrypsin by the pseudoirreversible inhibitor 2-ethoxy-4H-3,1-benz-oxazin-4-one. A substituted anthranilic acid [41470-93-3] (26) is formed.

Computer-Aided Inhibitor Design

Computer-aided inhibitor design is a relatively new and powerful approach for the development of novel, potentially potent, nonsubstrate-analogue enzyme inhibitors. Computer-aided methods and biological screening can each lead to new classes of novel inhibitors. However, computer-aided design methods can focus the search for inhibitors, thereby circumventing much of the time-consuming synthetic and natural product purification procedures for those compounds they find unlikely to function as inhibitors.

The development of a potent inhibitor must often be divided into two phases. First, a lead compound is developed; this compound is often defined somewhat arbitrarily as an enzyme inhibitor with a K_i value less than 10 μM , with a structure that can be identified in the enzyme-inhibitor complex, and with positions suitable for elaboration during lead optimization (37). Identification of a lead compound may be accomplished by biological screening, by any other inhibitor design method described elsewhere in this article, or by *de novo* design. The *de novo* design of inhibitors refers to creating from scratch a lead compound that is a new inhibitor of unique structure. Computer-aided *de novo* design methodologies currently require either the enzyme structure (eg, from x-ray crystallography) or the pharmacophore structure, which consists of the chemical groups of a ligand and their relative orientations important for binding to the enzyme. After obtaining a lead compound, its effectiveness may be optimized by using enzyme structural data as a starting point, or if there are many lead compounds of similar structure, inhibitor-analogue structural data and inhibitory activities may be used to begin the optimization process. In general, the data required for starting a computer-aided inhibitor design project are (1) the three-dimensional structure of the target enzyme or of a model constructed from related enzymes, or (2) the biological activities and structures of structurally related inhibitors of the particular enzyme, or (3) the pharmacophore for a particular enzyme, or a combination of the above. Comprehensive review articles on this subject are available (38–42).

DESIGN OF INHIBITORS FROM AN ENZYME STRUCTURE

As high quality, three-dimensional structures of enzymes have become increasingly available, they have been used for the structure-based design of inhibitors. Most of these enzyme structures have been solved by x-ray crystallography, because this method provides the required atomic-level resolution. The cavities forming the active site in the target enzyme structure are used to determine the shape of a potential inhibitor. By simply occupying space within the active site, an inhibitor obtains much of its binding energy, entropic in nature, from the hydrophobic effect (1,7,8). Therefore, maximally occupying the available space is favorable. Other sources of binding energy include van der Waals contacts (induced dipole interactions), hydrogen-bond formation, and positive and negative charge interactions (1,7,8). The enzyme structure provides a map showing the shape of available spaces, the locations of potential hydrogen-bonding groups, and the locations of potentially charged atoms.

Computers are integral to structure-based design processes. Graphics workstations and a variety of computer algorithms (procedures) are used to help visualization of three-dimensional structures. Molecular graphics programs can provide tools for displaying, building, and energy-minimizing three-dimensional structures, in addition to docking structures, bringing them together in three-dimensional space like an inhibitor into the active-site of an enzyme. Visual display requires molecular surface contours (43), color codes, depth perception cues, and tools for three-dimensional manipulation. Currently available program packages that contain molecular graphics capabilities include BIOGRAPH (from BioDesign), CHEM-X (from Chemical Design, Ltd.), FRODO (44,45), INSIGHT (from Biosym Technologies, Inc.) MACROMODEL (46,47), MIDAS (48,49), MOGLI (from Evans and Sutherland Computer Corp.) QUANTA (from Polygen Corp.) (50), and SYBYL (from Tripos Assoc.) (see MOLECULAR MODELING).

In addition to assisting with structure visualization, computer algorithms are used to automate the search for inhibitors by performing tasks such as screening databases and analyzing enzyme active sites. DOCK (51–54), a computer program package, brings molecules together in space to score the fit of a single potential inhibitor in an enzyme active site, or to screen small molecule databases for compounds that score well. The scores are based on shape complementarity, how snugly the test molecule fills the specified site, and electrostatic and energy force-field considerations. Available small molecule databases include the Cambridge Structural Database (CSD) (55), which contains x-ray crystal structures of close to 100,000 molecules, and the Fine Chemicals Directory (FCD) distributed by Molecular Design, Ltd (56), which contains approximately 55,000 computer-generated three-dimensional structures of commercially available compounds. Another program, GRID (57), finds regions in an active site that should have favorable interactions with a specific chemical group like a hydroxyl or methyl. The program maps this information as color-coded contours superimposed on the three-dimensional structure of the enzyme. A third program, HINT (Hydrophobic INTERactions) (58,59), identifies and maps potential hydrophobic and polar interactions between an inhibitor and an enzyme active site.

The greatest successes with structure-based design to date have been accomplished by utilizing x-ray crystal structures of enzymes complexed with bound inhibitors rather than uncomplexed enzyme structures. The structures of enzyme complexes are thought to capture the enzyme conformations that are especially favorable for ligand binding (37). Data of this nature may either be used to modify the inhibitor bound in the active site to improve its affinity for the enzyme or to search for novel inhibitors after removing the structure of the bound inhibitor from the displayed complex. The approach taken for structure-based design is shown schematically in Figure 6. With many iterations of this cycle, it is hoped that lead compounds can be developed and subsequently improved. It is also important to determine if the lead compound inhibits in the predicted manner or if inhibition occurs serendipitously.

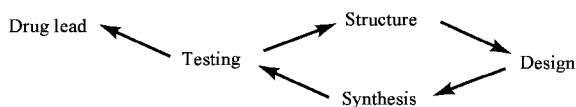


Fig. 6. The structure-based inhibitor design cycle.

An example of a structure-based methodology has been published (37) in which potent enzyme inhibitors are designed from templates at least somewhat different from those of natural substrates, cofactors, or currently existing inhibitors. In this study (37), the structure of thymidylate synthase (TS) from *E. coli* was used for the discovery of cancer-fighting lead compounds in humans. TS catalyzes the methylation of deoxyuridylate monophosphate [964-26-1] (27) to deoxythymidylate monophosphate [365-07-1] (29), while converting the cofactor, *N*⁵,*N*¹⁰-methylenetetrahydrofolate [3432-99-3] (28), into dihydrofolate [4033-27-6] (30) (Figure 7). The x-ray crystal structure of *E. coli* TS complexed with two substrate-analogue inhibitors, the covalently bound 5-fluorodeoxyuridylate and compound (31), was used for the *de novo* design of inhibitors. This ternary complex was used to generate the ligand-bound enzyme conformation, which differed significantly from that of the free enzyme. The inhibitor (31) was removed from the computer structure of the *E. coli* model for human TS so that the space could be used for developing novel inhibitors. By using this structural model and the program GRID (57), the binding energetics in the active site region previously occupied by (31) were analyzed, and complementary small molecule fragments were assembled into full-sized putative inhibitors. These compounds were synthesized and tested in enzyme assays, resulting in novel naphthostyryl-based and tetrahydroquinoline-based inhibitors (eg, (32) and (33) respectively) (Fig. 8) that were used as lead compounds for drug discovery. Each time an inhibitor was identified in the enzyme assay, the x-ray crystal structure of the new enzyme-inhibitor complex was solved. These new structures were used for subsequent design efforts to optimize the inhibitors' effectiveness. After several iterations of the structure-based design cycle, inhibitors of human TS with *K*_i values in the nanomolar (nM) range were developed. Follow-up work discussing the optimization of benz[*a*]indole-containing lead compounds has also been reported (60).

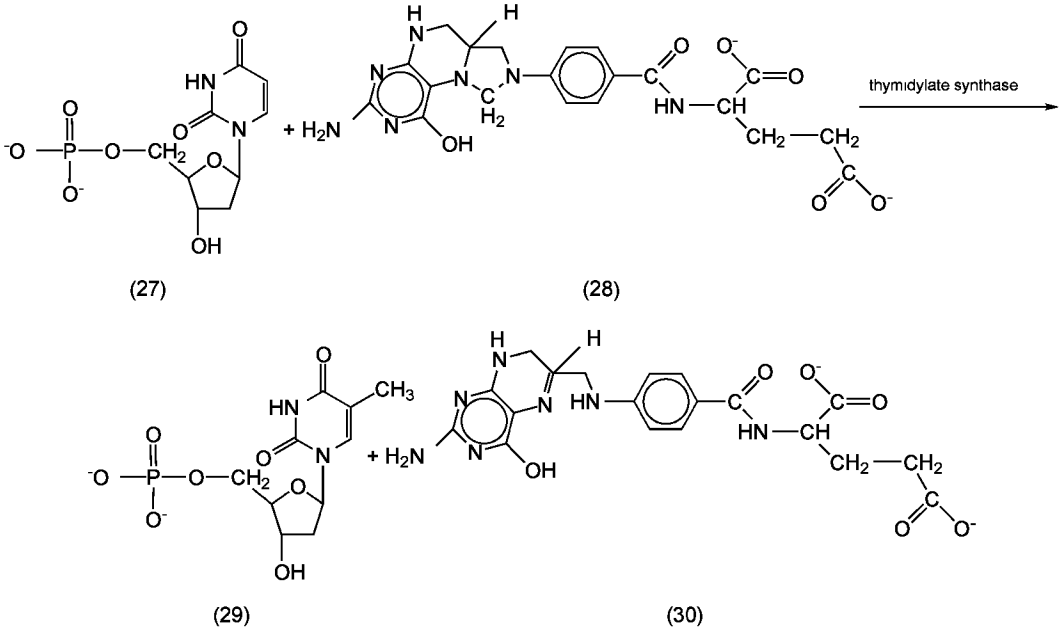


Fig. 7. Reaction catalyzed by thymidylate synthase.

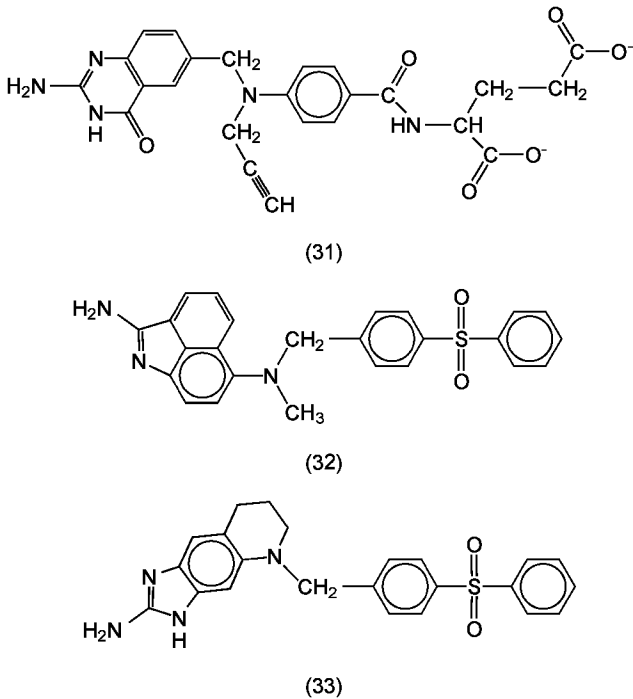


Fig. 8. Thymidylate synthase inhibitors designed to fit into the *N*⁵,*N*¹⁰-methylenetetrahydrofolate binding site. The best inhibitors from each class are the classical antifolate TS inhibitor (31), the naphthostyryl-based lead compound (32), and the tetrahydroquinoline-based lead compound (33), *K*_i values (in nM) for *E. coli* and human TS are as follows:

Structure	<i>E. Coli</i>	Human
(31)	3	12
(32)	1500	31
(33)	43	31

Another approach to the design phase of the structure-based design cycle utilizes the computer program DOCK, which increases the chances that compounds chosen for testing will be inhibitors. Novel lead compounds have been developed for *Lactobacillus casei* thymidylate synthase by using DOCK, x-ray crystallography, and additional computational analysis for the selection of potential inhibitors from a pool of commercially available compounds (61). Initial, weak-binding inhibitors were found by using DOCK and the x-ray crystal structure of TS to screen the small molecule database, FCD. X-ray crystal structures of resulting enzyme-inhibitor complexes were further analyzed, generating a pool of potential inhibitors. DOCK was then used to assess the complementarity of these molecules to the enzyme active site, and they were rank-ordered accordingly. As can be seen in Table 2, many of the top-scoring compounds showed inhibition in the μM range. The structures of the top ranked compounds with IC_{50} s at or below 100 μM are shown in Figure 9. These phenolphthalein derivatives are quite different in structure from the natural substrates of TS, which is useful for novel lead discovery.

Table 2. High-DOCK Scoring^a Ligands^b

Ligand	CAS Registry Number	Structure number	DOCK rank	IC ₅₀ , μM	DOCK score ^a
phenolphthalein	[77-09-8]	(34)	2	15	9.3
phenolphthymolphthalein		(35)	4	7	7.5
gibberellin A ₉			6	300	
2,2-bis(4-hydroxyphenyl)hexafluoropropane	[1478-61-1]	(36)	7	100	6.6
3,3-bis-(4-hydroxyphenyl)-oxindole	[125-13-3]	(37)	10	60	6.1

^a The best score is the most negative in value.
^b DOCK search of enzyme-inhibitor active site as described in text.

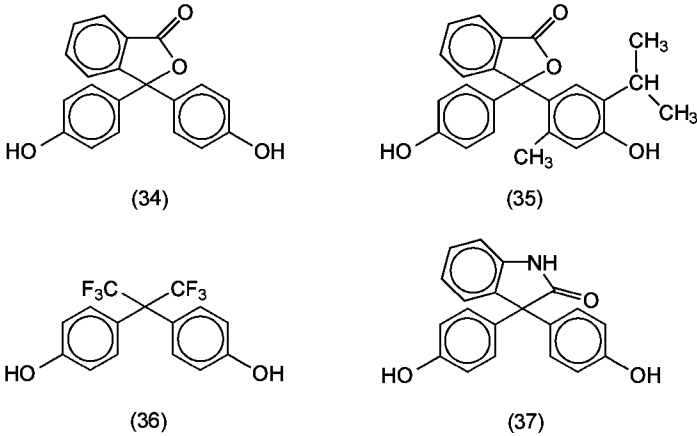


Fig. 9. Thymidylate synthase inhibitors found after analysis with the computer program DOCK (see Table 2).

DESIGN OF INHIBITORS FROM A SERIES OF STRUCTURALLY RELATED COMPOUNDS

Traditionally, medicinal chemists have used trial-and-error methods to study the relationship between structure and function for the development of inhibitors in a fashion more rational than biological screening. After measuring the inhibitory effect of an original compound, a small chemical modification is made to create a structural analogue, which is tested for a change in inhibition. It is assumed that the structural change is small enough so that the mechanism of action is not drastically altered. For example, after removing a methyl group from an inhibitor, the binding orientation of the analogue in the active site is assumed to remain the same as that for the original inhibitor, and any change in binding strength is attributed to the removal of the methyl group. In this manner, the medicinal chemist explores the spatial limits of the enzyme's binding site without knowing the structure of the enzyme. Modern, quantitative approaches use the speed and computational power of computer algorithms to determine specific structure-function relationships for the design of more potent analogues of known inhibitors.

Quantitative Structure-Activity Relationship (QSAR) Methods. These methods (62) are computer algorithms that construct mathematical functions correlating the chemical structure of a compound to its biological inhibitory activity. In many cases, this has led to the prediction of a compound's biological activity after obtaining data from a series of structurally similar molecules (63,64). For one QSAR approach, the Hansch method (65–67), it is assumed that the relative potency of an analogue is related to an additive combination of terms related to its physicochemical parameters such as pK_a , partition coefficient, and molecular refractivity. The values of the physicochemical properties and the values of concentration necessary for eliciting a standard biological effect are measured for a series of structurally related compounds. Next, a set of many equations with many unknowns is obtained and solved for by a least-squares multiple regression analysis. In another approach called the Free-Wilson method (68), it is assumed that the relative potency of an analogue is related to an additive combination of terms described by variables that indicate the presence or absence of functional groups. For systems that cannot be accurately described by either of these two methods alone, the methods can be combined. With hybrid Hansch–Free-Wilson equations, the effects of functional groups on biological activity that cannot be fully accounted for by the physicochemical properties alone can be quantified. Once the mathematical function has been determined by one of these methods, the variables for a new but related compound can be plugged into the equation, and the equations solved for its anticipated biological potency.

An important benefit of QSAR methods is that no enzyme structure is required. However, a series of structurally related inhibitors of known structure and known inhibitory activity is required. The limitations of the method involve the difficulties in describing something as complicated as a structure–function relationship in a single mathematical expression.

Comparative Molecular Field Analysis Methods. Comparative molecular field analysis (CoMFA) (69,70) has been referred to as 3D-QSAR. CoMFA methodologies are algorithms that relate the biological activities of a series of molecules to the steric (also known as van der Waals), and electrostatic energy fields calculated from their structural data. The CoMFA results are displayed three-dimensionally relative to the structures from which the energetics were originally calculated. As with QSAR, CoMFA has led to predictions of a molecule's biological activity after obtaining data from a series of structurally related compounds (71–73). For the CoMFA approach, a series of compounds that produces a specific biological response is identified, and three-dimensional structural models are constructed. These structures are then superimposed upon one another and placed within a fixed lattice. A probe atom, with its own energetic values, is systematically placed at each lattice point where it is used to calculate the steric and electrostatic potentials between itself and each of the superimposed structures. Again at each lattice point, these values are stored along with each inhibitor's biological

activity. Therefore, at each lattice point, one steric value, one electrostatic value, and one inhibition value is saved for each inhibitor in the series. Subsequently, for each inhibitor, one equation with many unknowns is established. For the entire series, the set of many equations with many unknowns is established, and each term's proportionality constant is solved for using the partial least squares method (69). The results of a CoMFA are represented as a three-dimensional coefficient contour map in which contours of various color represent locations on the structure where lower or higher steric and lower or higher electrostatic interactions would increase binding. This useful representation of the data allows a medicinal chemist to alter known structures and design potentially tighter binding compounds.

The biggest limitation of the CoMFA method is the alignment step. The algorithm superimposes the portions of the inhibitors that are of similar structure, assuming that they bind with similar orientations in the active site of the enzyme, which is not necessarily the case. Also, because of a problem with alignment, a CoMFA may fail when a few molecules are very dissimilar from all others in the series. Like QSAR, CoMFA does not require a structure of the relevant biological receptor, but does require knowledge about a series of inhibitory compounds.

Perturbation Free-Energy Calculations. These calculations directly calculate the effect of a small change in an inhibitor's structure on the strength of binding to the enzyme using cyclic thermodynamic descriptions of the Gibb's free energy changes. These methods require such a large amount of computer time that they are "currently not routine" (74). However, one example using this methodology has been reported for peptidomimetic inhibitors of the human immunodeficiency virus 1 protease (75).

DESIGN OF INHIBITORS FROM A PHARMACOPHORE

Pharmacophore-based inhibitor design is a method for developing inhibitors that combines the structural elements known to be important for binding with new structural templates. From a series of compounds that bind to the enzyme, the pharmacophore, the structural unit containing the chemical elements in the orientation required for binding, must first be identified. Examples of pharmacophores can be found in the literature (76–80). A pharmacophore for angiotensin converting enzyme has been proposed (79) (Fig. 10a). The key atoms for this pharmacophore are a carbonyl carbon, carbonyl oxygen, carboxylate oxygen, and an enzyme-bound zinc ion. A new pharmacophore model (80) (Fig. 10b) provides a more general structural description, thereby allowing for a broader database search, resulting in a larger variety of structures selected as hits that contain the pharmacophore. Also, since a zinc-containing pharmacophore was incompatible with the database used in their study, the second model included a zinc binding atom of the inhibitor rather than the zinc itself (80). Distance geometry, an area of mathematics used to build structures from internal distances (81,82), was used to generate ten conformations of the zinc complex consistent with the initial distance and angular description of the pharmacophore. The range for each distance and angle among all conformations was used to define the maximum and minimum boundaries for the distances and angles used to define the final pharmacophore model.

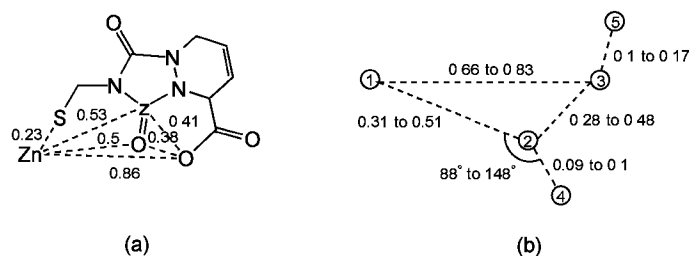


Fig. 10. Pharmacophores for angiotensin-converting enzyme. Distances in nm. (a) The structure of a semirigid inhibitor and distances between essential atoms from which one pharmacophore was derived (79). (b) In another pharmacophore, atom 1 is a potential zinc ligand (sulfhydryl or carboxylate oxygen), atom 2 is a neutral hydrogen bond acceptor, atom 3 is an anion (deprotonated sulfur or charged oxygen), atom 4 indicates the direction of a hydrogen bond to atom two, and atom 5 is the central atom of a carboxylate, sulfate, or phosphate of which atom 3 is an oxygen, or atom 5 is an unsaturated carbon when atom 3 is a deprotonated sulfur. The angle 1-2-3-4 is 135 to 180° or 135 to 180°, and 1-2-3-5 is 90 to 90°.

Once the pharmacophore has been defined, it can be used in many ways to develop inhibitors. For example, structural databases can be searched for existing molecules that contain the proposed pharmacophore (80). Alternatively, potential inhibitors can be built from structural pieces that connect the elements of the pharmacophore in the proper fashion (76). The pharmacophore for α -amylase was defined (76), and databases were searched to identify new low energy, bioactive templates (structural scaffolding) with conformations that could present the pharmacophore to the enzyme. After the templates were overlapped by superimposing common pharmacophore points, a synthetically feasible hybrid of the pharmacophore and a template was designed as a potential inhibitor of the target enzyme. Before the design was finalized, the hybrid was energy-minimized to verify that the pharmacophore remained intact in the molecule's low energy conformation. Some specific pharmacophore-based design algorithms include CAVEAT (76), Aladdin (83), and Spacer Skeletons (84,85). Inhibitors have been designed using pharmacophore-based methodologies (77,86).

Summary

A number of very powerful strategies are available to obtain potent enzyme inhibitors acting either in a covalent or noncovalent fashion. A great deal of progress has been made in the last 20 years towards the rational design of inhibitors, reducing significantly the need for laborious screening of a vast number of potential inhibitors for a particular target enzyme. Technological development has facilitated the rational design and the study of enzyme inhibitors, particularly through computer-aided drug design and genetic engineering. Recombinant DNA techniques enable the overexpression of proteins and potentially can provide large amounts of enzyme in comparison with traditional enzyme purification procedures. These advances facilitate the study of enzymes and their inhibitors enormously. Rapid development in the areas of x-ray crystallography and computer methods will continue to facilitate the discovery of lead compounds and their optimization to potent inhibitors, saving significant time and resources in the future.

BIBLIOGRAPHY

1. D. V. Santi and G. L. Kenyon in M. E. Wolff, ed., *Burger's Medicinal Chemistry*, 4th ed., Wiley-Interscience, New York, 1980, pp. 349–391.
2. G. R. Stark and P. A. Bartlett, *Pharmac. Ther.* **23**, 45–78 (1983).
3. P. Nuhn and T. Koehler, *Pharmacologie* **46**, 365–382 (1991).
4. R. B. Silverman, *The Organic Chemistry of Drug Design and Drug Action*, Academic Press, Inc., San Diego, 1992.
5. W. H. Frishman and R. C. Rapier in A. G. Gillman and co-workers, eds., *The Pharmacologic Basis of Therapeutics*, 7th ed., MacMillan, New York, 1985, p. 440.
6. I. H. Segel, *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*, Wiley-Interscience, New York, 1975.
7. A. Fersht, *Enzyme Structure and Mechanism*, 2nd ed., W. H. Freeman and Co., New York, 1985.
8. W. P. Jencks, *Adv. Enzymol. Relat. Areas Mol. Biol.* **43**, 219–410 (1975).
9. Q.-F. Ma and co-workers, *J. Med. Chem.* **35**, 1938–1941 (1992).
10. H. Yamashita, H. Nakatani, and B. Tonomura, *J. Biochem.* **111**, 182–185 (1992).
11. Z. Debyser and co-workers, *Proc. Natl. Acad. Sci. USA* **88**, 1451–1455 (1991).

12. J. F. Morrison, *Trends Biochem. Sci.* **7**, 102–105 (1982).
13. D. H. Rich, *J. Med. Chem.* **28**, 263–273 (1985).
14. J. F. Morrison and C. T. Walsh, *Adv. Enzymol. Relat. Areas Mol. Biol.* **61**, 201–301 (1988).
15. R. Wolfenden in R. D. Gandour and R. L. Schowen, eds., *Transition States of Biochemical Processes*, Plenum Press, New York, 1978, pp. 555–578.
16. R. Wolfenden, *Annu. Rev. Biophys. Bioeng.* **5**, 271–306 (1976).
17. G. Lienhard, *Science* **180**, 149–154 (1973).
18. P. Bartlett and co-workers, *J. Org. Chem.* **53**, 3195–3210 (1988).
19. P. A. Bartlett and C. K. Marlowe, *Biochemistry* **22**, 4618–4624 (1983).
20. R. H. Abeles, *Drug Dev. Res.* **10**, 221–234 (1987).
21. D. H. Kinder, S. K. Frank, and M. M. Ames, *J. Med. Chem.* **33**, 819–823 (1990).
22. A. D. Broom, *Fed. Proc.* **45**, 2779–2783 (1986).
23. A. D. Broom, *J. Med. Chem.* **32**, 2–7 (1989).
24. R. M. Lindquist in S. J. Ariens, ed., *Medicinal Chemistry*, Academic Press, Inc., New York, 1975, pp. 24–80.
25. J. Inglese, S. J. Blatchly, and S. J. Benkovic, *J. Med. Chem.* **32**, 937–940 (1989).
26. R. H. Abeles, *Chem. Eng. News* **61**, 48–56 (1983).
27. T. I. Kalman, *Drug Dev. Res.* **1**, 311–328 (1981).
28. M. G. Palfreyman, P. Bey, and A. Sjoerdsma, *Essays Biochem.* **23**, 28–81 (1987).
29. R. B. Silverman, *Mechanism-Based Enzyme Inactivation: Chemistry and Enzymology*, CRC Press, Boca Raton, Fla., 1988.
30. R. B. Silverman, *J. Enzym. Inhib.* **2**, 73–90 (1988).
31. C. T. Walsh, *Tetrahedron* **38**, 871–909 (1982).
32. C. T. Walsh, *Ann. Rev. Biochem.* **53**, 493–535 (1984).
33. N. A. Thornberry and co-workers, *J. Biol. Chem.* **266**, 21, 657–21, 665 (1991).
34. W. Jacoby and M. Wilchek, eds., *Methods in Enzymology, Affinity Labeling* **46**, Academic Press, Inc., New York, 1977.
35. D. Buechter and co-workers, *J. Biol. Chem.* **267**, 2173–2178 (1992).
36. L. Hedstrom and co-workers, *Biochemistry* **23**, 1753–1759 (1984).
37. K. Appelt and co-workers, *J. Med. Chem.* **34**, 1925–1934 (1991).
38. N. C. Cohen and co-workers, *J. Med. Chem.* **33**, 883–894 (1990).
39. C. Hutchins and J. Greer, *Crit. Rev. Biochem. Mol. Biol.* **26**, 77–127 (1991).
40. I. D. Kuntz, *Science* **257**, 1078–1082 (1992).
41. Y. C. Martin in K. Lipkowitz and D. Boyd, eds., *Reviews in Computational Chemistry*, VCH Press, New York, 1990, pp. 213–263.
42. Y. C. Martin, *Methods Enzymol.* **203**, 587–613 (1991).
43. M. L. Connolly, *Science* **211**, 709–713 (1983).
44. T. A. Jones, *J. Appl. Cryst.* **11**, 268–272 (1978).
45. T. A. Jones, *Computational Crystallography*, Clarendon Press, Oxford, 1982.
46. F. Mohamadi and N. G. J. Richards, *J. Comp. Chem.* **11**, 440–467 (1990).
47. W. C. Still and co-workers, *Tetrahedron* **40**, 2775–2281 (1984).
48. T. E. Ferrin and co-workers, *J. Mol. Graph.* **6**, 13–27 (1988).
49. *Ibid.*, pp. 2–12.
50. B. R. Brooks and co-workers, *J. Comp. Chem.* **4**, 187–217 (1983).
51. R. L. DesJardais and co-workers, *J. Med. Chem.* **31**, 722–729 (1988).
52. I. D. Kuntz and co-workers, *J. Mol. Biol.* **161**, 269–288 (1982).
53. E. C. Meng, B. K. Shoichet, and I. D. Kuntz, *J. Comp. Chem.* **13**, 505–524 (1992).
54. B. K. Shoichet, D. Bodian, and I. D. Kuntz, *J. Comp. Chem.* **13**, 380–397 (1992).
55. F. H. Allen and co-workers, *Acta. Crystallogr. B* **35**, 2331–2339 (1979).
56. O. F. Gunner, D. W. Hughes, and L. M. Dumont, *J. Chem. Inf. Comput. Sci.* **31**, 408–414 (1991).
57. P. J. Goodford, *J. Med. Chem.* **28**, 849–857 (1985).
58. G. E. Kellogg, S. F. Semus, and D. J. Abraham, *J. Comput. Aided Mol. Des.* **5**, 545–552 (1991).
59. F. C. Wireko, G. E. Kellogg, and D. J. Abraham, *J. Med. Chem.* **34**, 758–767 (1991).
60. M. D. Varney and co-workers, *J. Med. Chem.* **35**, 663–676 (1992).
61. B. K. Shoichet and co-workers, *Science* **259**, 1445–1450 (1993).
62. Y. C. Martin, *J. Med. Chem.* **24**, 229–237 (1981).
63. R. D. Cramer III and co-workers, *J. Med. Chem.* **22**, 714–725 (1979).
64. J. P. Dirdam and co-workers, *J. Med. Chem.* **22**, 1118–1121 (1979).
65. C. Hansch and T. Fujita, *J. Am. Chem. Soc.* **86**, 1616–1626 (1964).
66. C. Hansch, *Acc. Chem. Res.* **2**, 232–239 (1969).
67. C. Hansch, *J. Med. Chem.* **19**, 1–6 (1976).
68. S. M. Free Jr. and J. W. Wilson, *J. Med. Chem.* **7**, 395–399 (1964).
69. R. D. Cramer III, D. E. Patterson, and J. D. Bunce, *J. Am. Chem. Soc.* **110**, 5959–5967 (1988).
70. R. D. Cramer III, D. E. Patterson, and J. D. Bunce, *Prog. Clin. Biol. Res.* **29**, 161–165 (1989).
71. P. G. DeBenedetti and co-workers, *J. Med. Chem.* **30**, 459–464 (1987).
72. G. D. Diana and co-workers, *J. Med. Chem.* **35**, 1002–1008 (1992).
73. J. W. McFarland, *J. Med. Chem.* **35**, 2543–2550 (1992).
74. J. A. McCammon in C. A. Ramsden, ed., *Comprehensive Medicinal Chemistry*, Pergamon Press, Oxford, 1990, pp. 139–152.
75. M. R. Reddy, V. N. Viswanadhan, and J. N. Weinstein, *Proc. Natl. Acad. Sci. USA* **88**, 10,287–10,291 (1991).
76. P. A. Bartlett and co-workers in S. M. Roberts, S. V. Ley, and M. M. Campbell, eds., *Chemical and Biological Problems in Molecular Recognition*, Royal Society of Chemistry, London, 1989, pp. 182–196.
77. J. Butera and co-workers, *J. Med. Chem.* **32**, 757–765 (1989).
78. T. L. Macdonald and co-workers in M. Potmesil and K. W. Kohn, eds., *Topoisomerases in Cancer*, Oxford University Press, New York, 1991, pp. 199–214.
79. D. Mayer and co-workers, *J. Comput. Aided Mol. Des.* **1**, 3–16 (1987).
80. R. P. Sheridan and co-workers, *Proc. Natl. Acad. Sci. USA* **86**, 8165–8169 (1989).
81. L. M. Blumenthal, *Theory and Applications of Distance Geometry*, Chelsea, Bronx, New York, 1970.
82. I. D. Kuntz, J. F. Thomason, and C. M. Oshiro, *Methods Enzymol.* **177**, 159–205 (1989).
83. J. H. VanDrie, D. Weininger, and Y. C. Martin, *J. Comput. Aided Mol. Des.* **3**, 225–251 (1989).
84. R. A. Lewis and P. M. Dean, *Proc. R. Soc. Lond. B* **236**, 141–162 (1989).

85.

Ibid., pp. 2–12.

86.

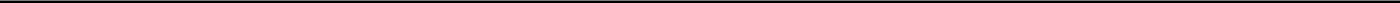
F. Leteurtre and co-workers, *Cancer Res.* **52**, 4478–4483 (1992).

Angelika Muscate

Cynthia L. Levinson

George L. Kenyon

University of California, San Francisco



ENZYMES, IMMOBILIZED.

See ENZYME APPLICATIONS; ENZYMES IN ORGANIC SYNTHESIS.

ENZYMES IN ORGANIC SYNTHESIS

The application of enzymes in organic synthesis is a modern and rapidly growing area in synthetic organic chemistry. Although the great potential of enzymatic catalysis had been recognized for many years, its application was rather limited. The rapid expansion of this area in recent years has been brought about by a number of factors, most importantly that a large number of enzymes have become commercially available. Of about 2500 enzymes identified thus far about 300 are available in a partly purified form. Moreover, because of advances in molecular biology, fermentation, and purification techniques, the cost of some enzymes has been reduced to less than \$500 kg.

The synthetic utility of enzymes for the preparation of organic molecules is tremendous. Enzymes catalyze virtually all types of chemical reactions with the exception of Diels-Alder condensation. They possess remarkably high catalytic power (up to 10¹² rate acceleration compared to the nonenzymatic reactions) and unsurpassed stereo- and regioselectivity. From a technological standpoint they also offer a number of advantages: the mild temperatures, neutral pH, and atmospheric pressure under which most enzymes operate result in processes that are environmentally acceptable, low energy consuming, and usually do not require high capital investments.

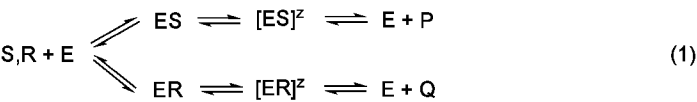
Biotransformations are carried out by either whole cells (microbial, plant, or animal) or by isolated enzymes. Both methods have advantages and disadvantages. In general, multistep transformations, such as hydroxylations of steroids, or the synthesis of amino acids, riboflavin, vitamins, and alkaloids that require the presence of several enzymes and cofactors are carried out by whole cells. Simple one- or two-step transformations, on the other hand, are usually carried out by isolated enzymes. Compared to fermentations, enzymatic reactions have a number of advantages including simple instrumentation; reduced side reactions, easy control, and product isolation.

Enzymatic transformations have been the subject of the most extensive investigation in recent years, resulting in the publication of thousands of original papers. The principal emphasis of this article is on reactions carried out by isolated enzymes that have the broadest synthetic utility, ie, hydrolases, oxidoreductases, and lyases. Biotransformations catalyzed by living cells are considered to a lesser extent. Enzymes that are used on a large scale such as amylase and glucose isomerase and also endonucleases and ligases that are routinely used for DNA synthesis are extensively covered elsewhere (see ENZYME APPLICATIONS, INDUSTRIAL; GENETIC ENGINEERING). For more detailed information on biotransformations the reader may consult several books and numerous reviews (1–14).

Hydrolases

Quantitative Analysis of Selectivity. One of the principal synthetic values of enzymes stems from their unique enantioselectivity, ie, ability to discriminate between enantiomers of a racemic pair. Detailed quantitative analysis of kinetic resolutions of enantiomers relating the extent of conversion of racemic substrate (*c*), enantiomeric excess (*ee*), and the enantiomeric ratio (*E*) has been described in an excellent series of articles (7,15,16).

During a resolution process, the R- and S-enantiomers compete for the free enzyme to form the noncovalent enzyme–substrate complexes ES and ER. These proceed to form transition-state intermediates [ES][‡] and [ER][‡]:



Since the rates for Michaelis-Menten kinetics at the steady state are described by

v_S

$\sim (k_{cat} / K_m)_S [E] [S]$

and

v_R

$\sim (k_{cat} / K_m)_R [E] [R]$

where *K_m* is the Michaelis constant, then

v_S

$\sim (k_{cat} / K_m)_S [S]$

v_R

$\sim (k_{cat} / K_m)_R [R]$

The ratio of specificity constants (*k_{cat}* / *K_m*)_S and (*k_{cat}* / *K_m*)_R determines the enantioselectivity of the reaction. Since (see Fig. 1)

$(k_{cat} / K_m)_S \sim \exp(\Delta G_S^\ddagger / RT)$

and

$(k_{cat} / K_m)_R \sim \exp(\Delta G_R^\ddagger / RT)$

where *R* and *T* are the gas constant and reaction temperature respectively, then

$(k_{cat} / K_m)_S / (k_{cat} / K_m)_R = \exp(\Delta \Delta G^\ddagger / RT) = E$

(2)

where *E* is the enantiomeric ratio that depends only on the specificity constant and is independent of time and substrate concentrations. Hence enantioselectivity of an enzyme is brought about by the free-energy difference of the diastereomeric transition states for the two enantiomers. Equation 2 can be converted into 3 that relates *E* to degree of conversion, *c*, and enantiomeric excess, *ee_S*:

$\ln ([1 - c] / [1 - ee_S]) = \ln ([1 - c] / [1 + ee_S]) = E$

(3)

Theoretical plots of *ee_S* (substrate) and *ee_P* (product) as a function of *c* are shown in Figure 2a and b. It can be seen that the *ee_S* increases with the extent of conversion. Consequently the enantiomeric purity of the substrate can be increased by sacrificing the yield and carrying out the reaction to higher degrees of conversion. Conversely, if high purity product is required the conversion should be terminated at early stages.

Kirk-Othmer Encyclopedia of Chemical Technology (4th Edition)

331

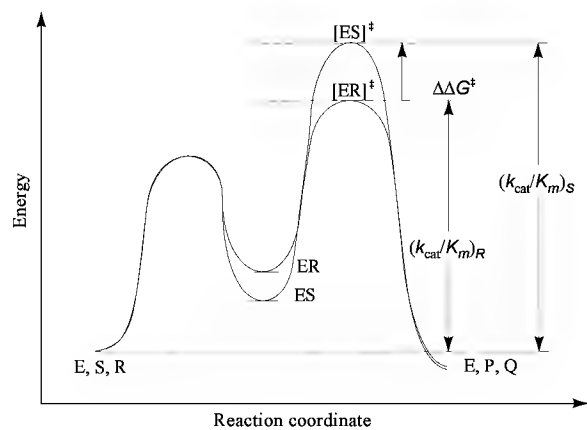


Fig. 1. Free-energy profile for a kinetic resolution depicted by equation 1 that follows Michaelis-Menten kinetics.

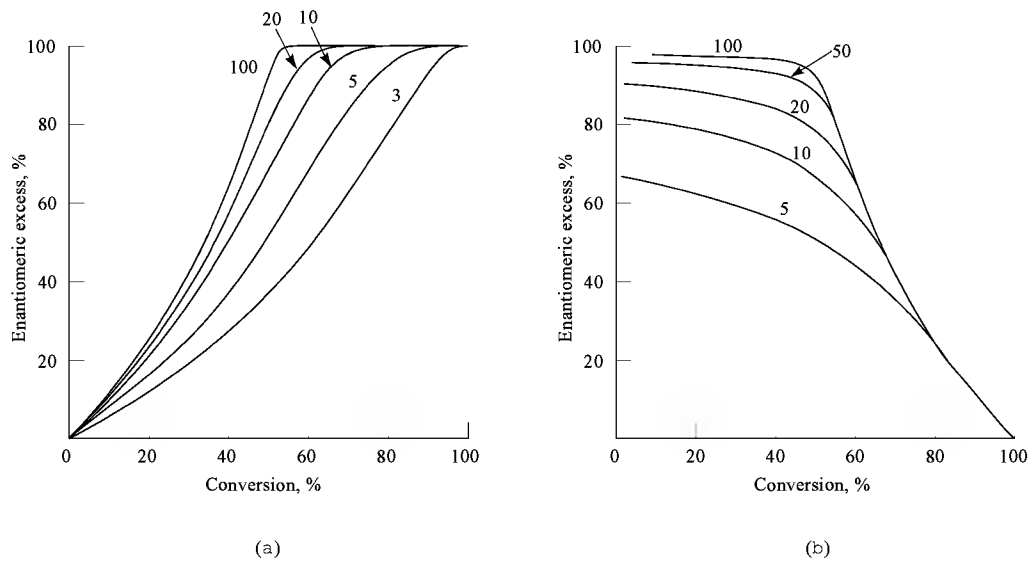


Fig. 2. Enantiomeric excess (*ee*) as a function of the conversion for various enantiomeric ratios (*E*) noted on the curves: (a) remaining substrate, $ee_S = (R - S) / (R + S)$; (b) product $ee_P = (P - Q) / (P + Q)$. Reprinted with permission (15).

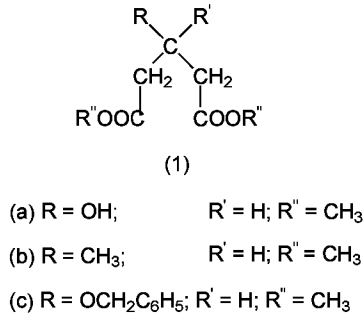
Courtesy of the American Chemical Society.

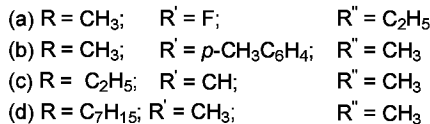
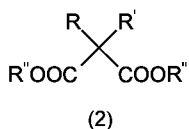
Enzyme-Catalyzed Asymmetric Synthesis. The extent of kinetic resolution of racemates is determined by differences in the reaction rates for the two enantiomers. At the end of the reaction the faster reacting enantiomer is transformed, leaving the slower reacting enantiomer unchanged. It is apparent that the maximum product yield of any kinetic resolution cannot exceed 50%. The situation is different if the substrate is a prochiral or meso compound. Since these molecules have a center or plane of symmetry the binding of pro-S or pro-R forms is equivalent. The chirality appears only as a result of the transformation. Hence, at least theoretically, the compound can be converted to one enantiomer quantitatively.

Hydrolytic enzymes such as esterases and lipases have proven particularly useful for asymmetric synthesis because of their abilities to discriminate between enantiotopic ester and hydroxyl groups. A large number of esterases and lipases are commercially available in large quantities; many are inexpensive and accept a broad range of substrates.

Dicarboxylic Acid Monoesters. Enzymatic synthesis of monoesters of dicarboxylic acids by hydrolysis of the corresponding diesters is a widely used and thoroughly studied reaction. It is catalyzed by a number of esterases, lipases, and proteases and is usually carried out in an aqueous buffer, pH 6–8 at room temperature. Organic cosolvents may be added to increase solubility of the substrates. The pH is maintained at a constant level by the addition of aqueous hydroxide. After one equivalent of base is consumed the monoesters are isolated by conventional means.

A number of examples of enantioselective hydrolysis of diesters (1 and 2) of malonic and glutaric acids are given in Table 1.





Porcine liver esterase (PLE) gives excellent enantioselectivity with both dimethyl 3-methylglutarate [19013-37-7] (1b) and malonate (2b) diester. It is apparent from Table 1 that the enzyme's selectivity strongly depends on the size of the alkyl group in the 2-position. The hydrolysis of ethyl derivative (2c) gives the S-enantiomer with 75° *ee* whereas the hydrolysis of heptyl derivative (2d) results in the R-monoester with 90° *ee*. Chymotrypsin [9004-07-3] (CT) does not discriminate glutarates that have small substituents in the 3-position well. However, when hydroxyl is replaced by the much bulkier benzyl derivative (1c), enantioselectivity improves significantly.

Table 1. Enantioselective Hydrolysis of Malonate^a and Glutarate^b Esters

Substrate	Enzyme	Monoester			References
		Yield, %	<i>ee</i> , %	Configuration	
(1a)	CT	78	55 ± 5	R	17
	PLE	76	30 ± 5	S	17
(1b)	PLE	92	100	R	18
(1c)	CT	86	92	R	19
(2a)	lipase MY ^c	75	86	S	20
(2b)	PLE	^d	96	R	21
(2c)	PLE	90–98	75	S	22
(2d)	PLE	90–98	90	R	22

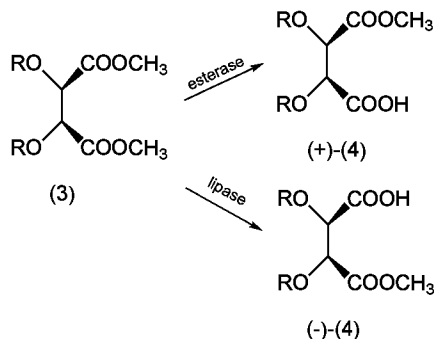
^a Structure (2).

^b Structure (1).

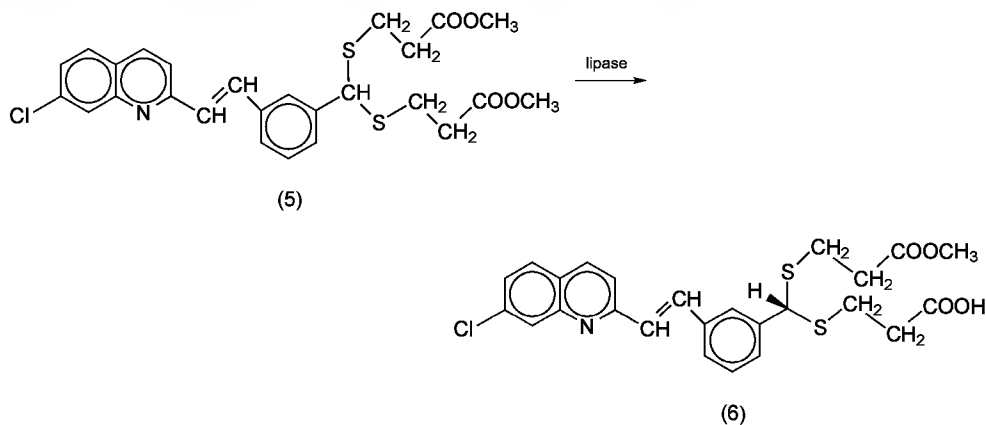
^c From *Candida cylindracea*.

^d Not reported.

Similar to the hydrolysis of malonates and glutarates, conversion of mesoderivatives of tartaric acid can be accomplished with *Candida cylindracea* lipase (CCL) and porcine liver esterase in very good yield (23). Hydrolysis of (3) with esterase gives (±)-(4) with an *ee* 90.5° in >90% yield. Hydrolysis with lipase results in the formation of (±)-(4) (*ee* 92.5°, >90% yield).

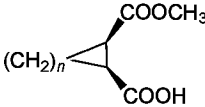
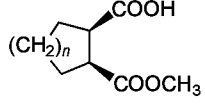
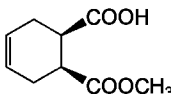
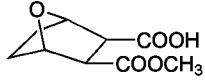
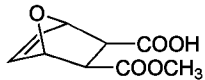
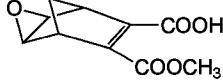
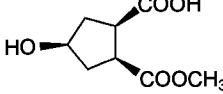
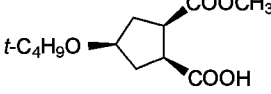
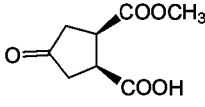
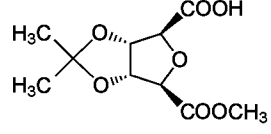


It is generally believed that selectivity of hydrolytic enzymes strongly depends on the proximity of the chiral center to the reacting carbonyl group, and only a few examples of successful resolutions exist for compounds that have the chiral center removed by more than three bonds. A noticeable exception to this rule is the enantioselective hydrolysis by *Pseudomonas fluorescens* lipase (PFL) of racemic dithioacetals (5) that has a prochiral center four bonds away from the reactive carboxylate (24). The monoester (6) is obtained in 89% yield and 98% *ee*.



PLE catalyzes the hydrolysis of a wide range of meso-diester (Table 2). This reaction is interesting from both theoretical and practical standpoints. Indeed, the analysis of a large range of kinetic data provided sufficient information to create a detailed active site model of PLE (31). From a practical standpoint, selective hydrolysis of *meso*-cyclo-1,2-dicarboxylates leads to chiral synthons that are valuable intermediates for the synthesis of a variety of natural products.

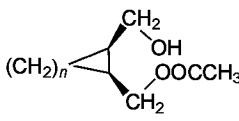
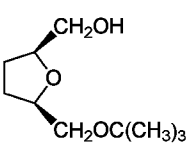
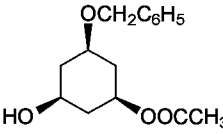
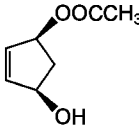
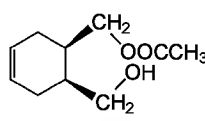
Table 2. PLE-Catalyzed Hydrolysis of Monocyclic meso-Diesters

Product	Yield, %	ee, %	References
 $n = 1,2$	87–88	>97	25
 $n = 1,2$	92($n = 1$) 88($n = 2$)	17 97	25
	98	98	26
	82	98	27
	86	75	27
	100	85	28
	76	80	29
	76	84	29
	86	88	29
	86	72	30

The hydrolysis of three- and four-membered rings (Table 2) is enantiotopically specific for the pro-(*S*)-methoxycarbonyl group leading to the formation of 1(*R*),2(*S*)-7 and 1(*R*),2(*S*)-8; in the case of five- and six-membered rings (9,10), the pro-(*R*)-group is hydrolyzed (25). This reversal of configurations of products was rationalized by postulating the presence of two binding cavities: large and small (31). Smaller molecules such as cyclopropane- and cyclobutanecarboxylates are accommodated in a smaller cavity leading to the formation of 1(*R*),2(*S*)-compounds. The larger rings can only bind in the larger cavity, thus reversing product configuration.

Hydrolysis of (11) that has a structure similar to (10) is also enantiotopically specific for the pro-(*S*)-methoxycarbonyl group and proceeds with very high yield (26). Bicyclic exo- (12,13) and endo-diesters that are valuable intermediates for the synthesis of natural cyclic compounds are excellent substrates for porcine pancreatic lipase (PPL) hydrolysis affording the corresponding monoesters in high chemical yield and optical purity (27). Dimethyl esters of meso-1,2-cyclopentanedicarboxylic acids (15–17) are also good substrates for PLE (29). Interestingly, the enzyme selectivity strongly depends on the nature of the substituent in the 4-position; whereas pro-(*R*)-ester group is preferentially hydrolyzed in hydroxyl diester (15), it is the pro-(*S*)-ester group that is cleaved in *tert*-butoxy diester (16). Hydrolysis of meso diesters of derivatives of 3,4-(isopropylidenedioxy)tetrahydrofuran (18) proceeds with only moderate selectivity giving monoesters with only 72% ee (32).

The transformations described thus far were catalyzed by enzymes in their traditional hydrolytic mode. More recent developments in the area of enzymatic catalysis in nonaqueous media (11,16,33–35) have significantly broadened the repertoire of hydrolytic enzymes. The acyl–enzyme intermediate formed in the first step of the reaction via acylation of the enzyme's active site nucleophile can be deacylated in the absence of water by a number of

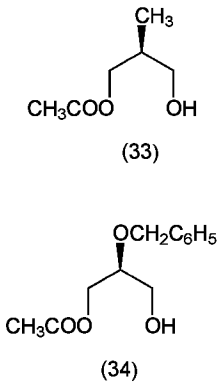
 $n = 1-4$	lipase <i>Mucor javanicus</i>	75	>99	44
	PLE	62	87	45
	PPL	89	100	46
	PPL	95	>99	47
				

^a PPL = porcine pancreatic lipase; PLE = porcine liver esterase.

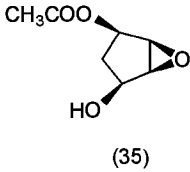
Cyclic diesters are often even better substrates for lipases and esterases than acyclic derivatives. Small-ring monoacetates (28, $n = 1-3$) are obtained in higher yield and ee than the larger derivatives (for 28, $n = 4$ ee is only 50% ee) (43). Hydrolysis of tetrahydrofuran diester results in monoester (29) of $ee > 99\%$ (44).

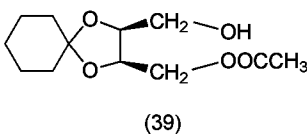
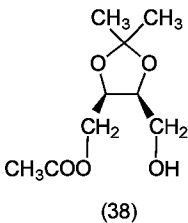
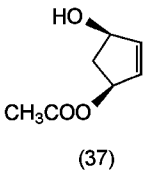
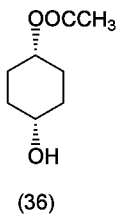
The synthetic utility of the above transformations stems from the fact that many monoesters obtained as a result of hydrolysis may be converted to pharmaceutically important intermediates. For example, the optically active glycerol derivative (27) is a key intermediate in the production of β -blockers. Allyl derivative (25) may be converted into (*S*)-paraconic acid [4694-66-0] ((*S*)-5-oxo-3-tetrahydrofurancarboxylic acid) that is a starting material for the synthesis of (3R)-A-factor. The unsaturated chiral cyclic monoacetate (31) is an optically active synthon for prostaglandins, and the monoester (29) is used for the synthesis of platelet activating factor (PAF) antagonists.

In contrast to the hydrolysis of prochiral esters performed in aqueous solutions, the enzymatic acylation of prochiral diols is usually carried out in an inert organic solvent such as hexane, ether, toluene, or ethyl acetate. In order to increase the reaction rate and the degree of conversion, activated esters such as vinyl carboxylates are often used as acylating agents. The vinyl alcohol formed as a result of transesterification tautomerizes to acetaldehyde, making the reaction practically irreversible. The presence of a bulky substituent in the 2-position helps the enzyme to discriminate between enantiotopic faces; as a result the enzymatic acylation of prochiral 2-benzyloxy-1,3-propanediol (34) proceeds with excellent selectivity ($ee > 96\%$) (49). In the case of the 2-methyl substituted diol (33) the selectivity is only moderate (50).



Enzymatic transesterification of cyclic diols often proceeds with very good yield and selectivity (35–39). For example, lipase PS-catalyzed acylation of the corresponding epoxy diol results in monoacetate (35) with $ee > 95\%$ (51). The same enzyme suspended in isopropenyl acetate which serves as both the reaction medium and the acylating agent catalyzes the conversion of the corresponding prochiral *cis*-2-cyclohexene-1,4-diol [53762-85-9] into monoacetate (36) in 98% ee (52). A crude mixture of enzymes obtained from pancreas, pancreatin, acylates 2-cyclopentene-1,4-diol [4157-01-1] to enantiomerically pure 1(*S*),4(*R*)-4-hydroxycyclopent-2-enyl acetate (37) (53). Homochiral monoester (38), which is a chiral building block for the synthesis of arachidonic acid metabolites, is obtained in 80% yield and 95% ee via lipase SAM-II-catalyzed transesterification of the corresponding diol with vinyl acetate (54). Lipase from *Pseudomonas cepacia* enantioselectively acylates (39) with acetic anhydride in toluene resulting in the product of R-configuration in 78% yield and >99% ee (55).





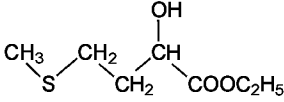
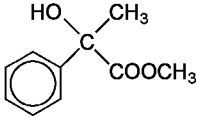
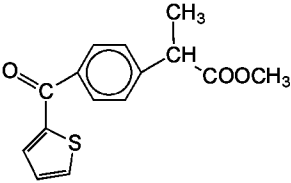
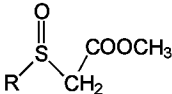
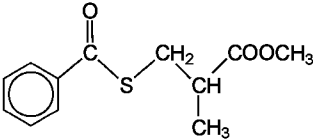
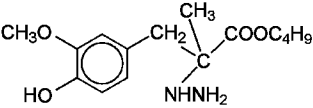
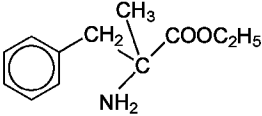
Kinetic Resolutions. From a practical standpoint the principal difference between formation of a chiral molecule by kinetic resolution of a racemate and formation by asymmetric synthesis is that in the former case the maximum theoretical yield of the chiral product is 50% based on a racemic starting material. In the latter case a maximum yield of 100% is possible. If the reactivity of two enantiomers is substantially different the reaction virtually stops at 50% conversion, and enantiomerically pure substrate and product may be obtained in close to 50% yield. Conveniently, the enantiomeric purity of the substrate and the product depends strongly on the degree of conversion so that even in those instances where reactivity of enantiomers is not substantially different, a high purity material may be obtained by sacrificing the overall yield.

The variety of enzyme-catalyzed kinetic resolutions of enantiomers reported in recent years is enormous. Similar to asymmetric synthesis, enantioselective resolutions are carried out in either hydrolytic or esterification–transesterification modes. Both modes have advantages and disadvantages. Hydrolytic resolutions that are carried out in a predominantly aqueous medium are usually faster and, as a consequence, require smaller quantities of enzymes. On the other hand, esterifications in organic solvents are experimentally simpler procedures, allowing easy product isolation and reuse of the enzyme without immobilization.

Optically Active Acids and Esters. Enantioselective hydrolysis of esters of simple alcohols is a common method for the production of pure enantiomers of esters or the corresponding acids. Several representative examples are summarized in Table 4. Lipases, esterases, and proteases accept a wide variety of esters and convert them to the corresponding acids, often in a highly enantioselective manner. For example, the hydrolysis of (*R*)-methyl hydratropate [34083-55-1] (40) catalyzed by lipase P from Amano results in the corresponding acid in 50% yield and 95% *ee* (56). Various substituents on the α -carbon (41–44) are readily tolerated by both lipases and proteases without reduction in selectivity (57–60). The enantioselectivity of many lipases is not significantly affected by changes in the alcohol component. As a result, activated esters may be used as a means of enhancing the reaction rate.

Table 4. Optically Active Acids and Esters Produced by Hydrolysis

Substrate	Enzyme	Ester			Acid			Refs
		Yield, %	<i>ee</i> , %	Configuration	Yield, %	<i>ee</i> , %	Configuration	
	lipase P	44	98	R	50	95	S	56
	lipase P-30	43	99	R	33	92	S	57
	α -chymotrypsin	46.5	80	S	38.5	75	R	58
	lipase P	40	>98	R	42	98	S	59

	protease <i>Asp. oryzae</i>	48	70	<i>S</i>	48	75	R	60
	lipase <i>C. cylindracea</i>	49 ^a	90	R	49 ^a	95	<i>S</i>	61
	lipase <i>Pseudomonas</i> sp.	33–49	98	<i>S</i>	17–38	88–98	R	62
 R = C ₆ H ₅ , <i>p</i> -ClC ₆ H ₄ , <i>p</i> -CH ₃ OC ₆ H ₄	lipase <i>A. niger</i>	32 ^a	45	<i>S</i>	45 ^a	98	R	63
	esterase <i>C. lipolytica</i>	50 ^a	98	R	50 ^a	>99	<i>S</i>	64
	esterase <i>C. lipolytica</i>	50 ^a	>99	R	^b	^b	<i>S</i>	64
								

^a Conversion.
^b Not reported.

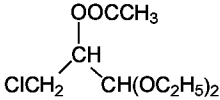
Lipase-catalyzed kinetic resolutions are often practical for the preparation of optically active pharmaceuticals (61). For example, suprofen [40828-46-4] (45), which is a nonsteroidal antiinflammatory drug, can be resolved by *Candida cylindracea* lipase in >95% *ee* at 49° o conversion (61). Moreover, lipase-based processes for the resolution of naproxen [22204-53-1] and ibuprofen [15687-27-1] (61) have also been developed.

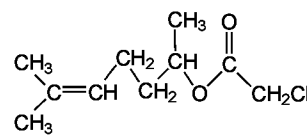
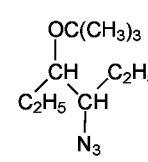
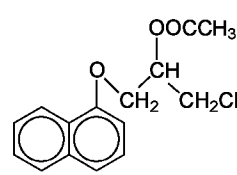
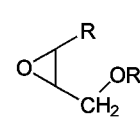
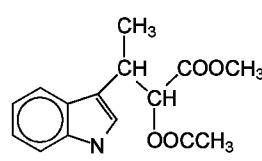
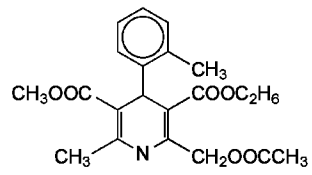
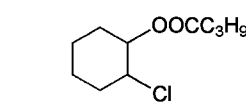
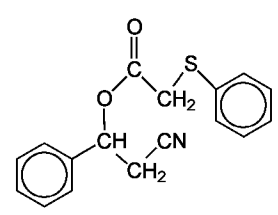
PPL and lipase from *Pseudomonas* sp. catalyze enantioselective hydrolysis of sulfinylalkanoates. For example, methyl sulfinylacetate (46) was resolved by *Pseudomonas* sp. lipase in good yield and excellent selectivity (62). This procedure was suitable for the preparation of sulfinylalkanoates where the ester and sulfoxide groups are separated by one or two methylene units. Compounds with three methylene groups were not substrates for the lipase (65).

The resolving of a variety of α-substituted carboxylic acid esters by a previously undescribed enzyme, *Candida lipolytica* esterase, has been reported (64). α-Methyl-α-amino (49) and α-methyl-α-hydrazino (48) esters, which function as inhibitors of acid decarboxylase enzymes, are obtained on a multigram scale in optically pure form.

Optically Active Alcohols and Esters. In addition to the hydrolysis of esters formed by simple alcohols described above, lipases and esterases also catalyze the hydrolysis of a wide range of esters based on more complex and synthetically useful cyclic and acyclic alcohols (Table 5). Although the hydrolysis of acetates often gives the desirable resolution, to achieve maximum selectivity and reaction efficiency, comparison of various esters is recommended.

Table 5. Optically Active Alcohols and Esters Produced by Hydrolysis

Substrate	Enzyme	Ester			Alcohol			Refs.
		Yield, ° o	<i>ee</i> , ° o	Config-uration	Yield, ° o	<i>ee</i> , ° o	Config-uration	
	lipase LP-80	46	>98	<i>S</i>	51.5	98	R	63
	lipase P-30	53	89	R	45	100	<i>S</i>	66

								
	lipase P	39	>98	R,S	35	>98	S,R	67
	lipase P	48	95	S	45	>95	R	68
 R = alkyl R' = acyl	PPL	32–45	90–95	2S,3R	15–40	30–65	2R,3S	69,30
	lipase OF-360	62	54	2R,3S	37	93	2S,3R	70
	lipase <i>Pseudomonas</i> sp.	50	91	R	50	98	R	71
	lipase <i>Pseudomonas</i> sp.	37	>98	R,R	46	>98	S,S	72
	lipase P	94	43	R	98	35	S	73

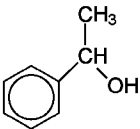
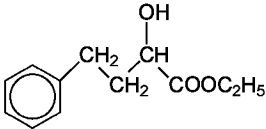
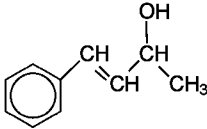
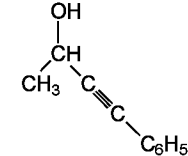
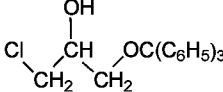
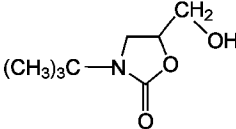
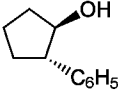
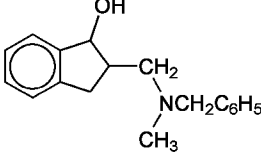
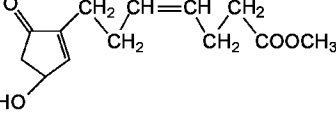
Both saturated (50) and unsaturated derivatives (51) are easily accepted by lipases and esterases. Lipase P from Amano resolves azide (52) or naphthyl (53) derivatives with good yields and excellent selectivity. PPL-catalyzed resolution of glycidyl esters (54) is of great synthetic utility because it provides an alternative to the Sharpless epoxidation route for the synthesis of β -blockers. The optical purity of glycidyl esters strongly depends on the structure of the acyl moiety; the hydrolysis of propyl and butyl derivatives of epoxy alcohols results in esters with $ee > 95\%$ (30).

Regioselective hydrolysis of diesters is a challenging problem in synthetic chemistry because the side reactions always reduce the yield of desired product. Some lipases are well suited to perform this task. Lipase OF-360 (Meito Sangyo) hydrolyzes diester (55) in 74° theoretical yield and 93° ee (70). Lipase from *Pseudomonas cepacia* suspended in diisopropyl ether saturated with water hydrolyzes triester (56) with a remarkable efficiency and regio- and stereoselectivity (71).

Two more examples in Table 5 include the hydrolysis of esters of trans-alcohols that proceed with high efficiency practically regardless of the nature of the substituents (72) and resolution of β -hydroxynitriles with lipase from *Pseudomonas* sp. In the latter case the enantioselectivity of the hydrolysis was improved by introducing sulfur into the acyl moiety (73).

Resolution of racemic alcohols by acylation (Table 6) is as popular as that by hydrolysis. Because of the simplicity of reactions in nonaqueous media, acylation routes are often preferred. As in hydrolytic reactions, selectivity of esterification may depend on the structure of the acylating agent. Whereas *Candida cylindracea* lipase-catalyzed acylation of racemic- α -methylbenzyl alcohol [98-85-1] (59) with butyric acid has an enantiomeric value E of 20, acylation with dodecanoic acid increases the E value to 46 (16). Not only acids but also anhydrides are used as acylating agents. *Pseudomonas fl* lipase (PFL), for example, catalyzed acylation of α -phenethanol [98-85-1] (59) with acetic anhydride in 42° yield and 92° selectivity (74).

Table 6. Optically Active Alcohols and Esters Produced by Acylation

Substrate	Enzyme	Alcohol			Ester			Refs
		Yield, %	ee, %	Configuration	Yield, %	ee, %	Configuration	
	lipase <i>P. fragi</i>	46	95	R	42	92	S	74
	PFL	48	98	S	51	98	R	75
	PPL	47 ^a	96	S	47 ^a	85	R	76
	lipase <i>Pseudomonas</i> sp.	47	>95	S	48	>95	R	77
	lipase PS	54	72	R	43	>98	S	78
	lipase <i>P. fluorescens</i>	42	95	S	40	95	R	74
	lipase <i>P. fluorescens</i>	43	95	S,R	46	95	R,S	79
	lipase P	46	98	S	43	97	R	80
	PPL	35	99.4	S	43	92	R	81

^a Conversion.

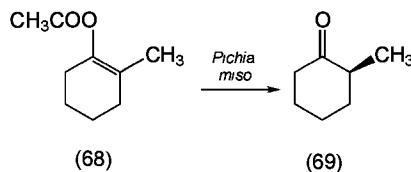
Various racemic secondary alcohols with different substituents, eg, α -hydroxyester (60), are resolved by PFL nearly quantitatively (75). The effect of adjacent unsaturation on enzyme-catalyzed kinetic resolutions was thoroughly studied for a series of allylic (61), propargylic (62), and phenyl-substituted 2-alkanols (76,77). Excellent selectivity was observed for (*E*)-allylic alcohols whereas (*Z*)-isomers showed poor selectivity (76).

Lipase-catalyzed enantioselective transesterification of *O*-substituted-1,2-diols is another practical route for the synthesis of β -blockers. Lipase PS suspended in toluene catalyzes the transesterification of (63) with vinyl acetate to give the (*S*)-ester in 43% yield and >98% *ee* (78). The desired product, optically pure (*R*)-tritylglycidol, is then easily obtained by treating the ester with alcoholic alkali. Moreover, *Pseudomonas* lipase catalyzes the acylation of oxazolidinone (64) with acetic anhydride in very good yield and selectivity (74). PPL-catalyzed transesterification of a number of *trans*-norbornene derivatives proceeds in about 30% yield and 92% *ee* (79,80).

Cyclic alcohols are excellent targets for enantioselective enzymatic acylations. For example, acylation of (65) with vinyl acetate catalyzed by lipase SAM-II gives the (*R*),(*S*)-ester with 95% *ee* (81). Similarly (66), which is a precursor for serotonin uptake inhibitor, is resolved in a high yield and selectivity with Amano lipase P (82). The prostaglandin synthon (67) is resolved by the same method into the optically pure alcohol in 35% yield (83).

Hydrolysis of Enol Esters. Enzyme-mediated enantioface-differentiating hydrolysis of enol esters is an original method for generating optically active α -substituted ketones (84–86). If the protonation of a double bond occurs from one side with the simultaneous elimination of the acyl group (Fig. 3), then the optically active ketone should be produced. Indeed, the incubation of 1-acetoxy-2-methylcyclohexene [1196-73-2] (68) with *Pichia*

miso affords (*S*)-2-methylcyclohexanone [22554-27-4] (69) in 77% yield and 90% *ee* (84).



Moreover, fermentation of various α -substituted cycloalkanone enol esters results in optically active six-, eight-, ten-, and twelve-membered ring ketones with 70–96% *ee* (84). Isolated enzymes catalyze similar transformations. *Bacillus coagulans* and *Candida cylindracea* lipase OF (Meito Sangyo) hydrolyze a number of cyclic and acyclic enol esters, giving ketones in 40–80% yield and 14–85% *ee* (85,86).

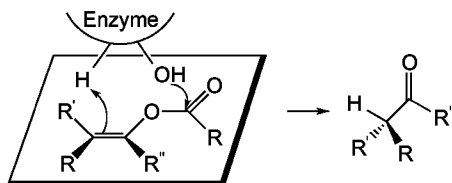
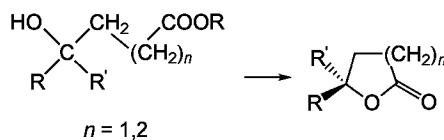
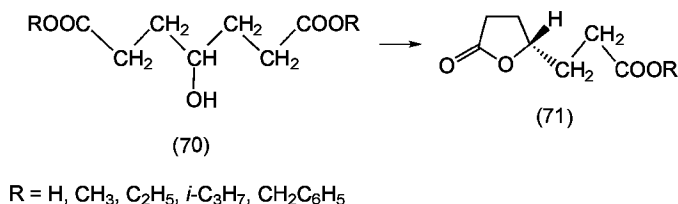


Fig. 3. Enzyme-mediated hydrolysis of an enol ester.

Chiral Lactones and Polyesters. Similar to intermolecular reactions described previously, lipases also catalyze intramolecular acylations of hydroxy acids; the reactions result in the formation of lactones.



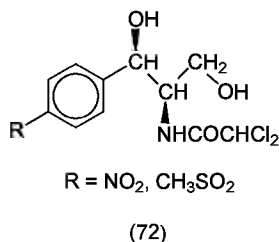
PPL suspended in dry ether catalyzes the lactonization of a number of γ -hydroxy acids. For example, (*S*)- γ -methylbutyrolactone [19041-15-7] ($R = \text{CH}_3$, $R' = \text{H}$, $n = 1$), and γ -phenylbutyrolactone ($R = \text{C}_6\text{H}_5$, $R' = \text{H}$, $n = 1$) may be produced in nearly quantitative yields (87–89). When prochiral derivatives (70) of γ -hydroxypimelic acid were submitted to the action of PPL and *Pseudomonas fluorescens* lipases, optically active lactones with up to 98% *ee* were produced.



The nature of the product strongly depends on the length of the hydroxy acid; generally when the hydroxyl group is remote the yield of lactone drops significantly. For example, 10-hydroxydecanoic acid [1679-53-4] does not produce any decanolide; instead, the reaction proceeds by intermolecular oligomerization, and a complex mixture of di-, tri-, tetra-, and pentalactones results (90). However, when *Pseudomonas* sp. or *Candida cylindracea* lipases are incubated with 16-hydroxyhexadecanoic acid [506-13-8], hexadecanolide is the predominant product (91).

Lipase-catalyzed intermolecular condensation of diacids with diols results in a mixture of macrocyclic lactones and linear oligomers. Interestingly, the reaction temperature has a strong effect on the product distribution. The condensation of α,ω -diacids with α,ω -diols catalyzed by *Candida cylindracea* or *Pseudomonas* sp. lipases leads to macrocyclic lactones at temperatures between 55 and 75°C (91), but at lower temperatures (<45°C) the formation of oligomeric esters predominates. Optically active trimers and pentamers can be produced at room temperature by PPL or *Chromobacterium viscosum* lipase-catalyzed condensation of bis(2,2,2-trichloroethyl) ($\pm$)-3-methyladipate and 1,6-hexanediol (92).

Regioselective Acylation of Hydroxy Compounds. Aliphatic diols can be selectively acylated at the primary position by a number of lipases in nonaqueous solvents. For example, PPL suspended in solutions of various diols in ethyl carboxylates catalyzes transesterification in a highly regioselective manner, producing primary monoesters in up to 97% yield (93). Similarly, chloramphenicol [56-75-7] (72) ($R = \text{NO}_2$) can be acylated by a number of lipases to produce optically pure, water-insoluble 3-*O*-palmitate in a highly selective manner (94).



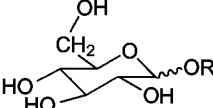
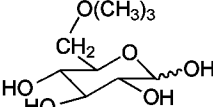
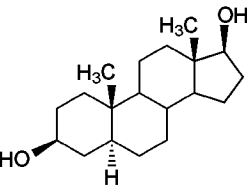
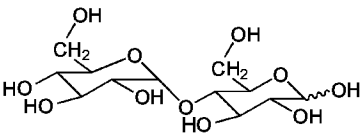
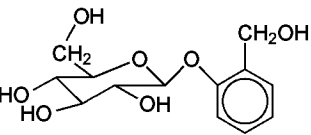
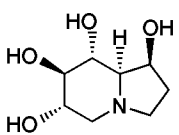
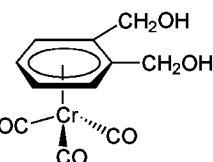
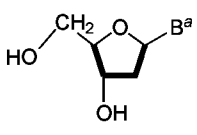
Acid anhydrides are also useful for acylation of primary alcohols in organic solvents. The reaction produces high yields of the primary acylated products with only traces (1–2%) of diesters (95).

This preference for the primary position is preserved during the acylation of polyhydroxy compounds. For example, PPL suspended in pyridine acylates glucose (73), galactose, mannose, and fructose with trichloroethyl carboxylates in a regioselective manner with an overwhelming preference toward C-6 primary hydroxyl (96) (Table 7). Similarly, the primary hydroxyl of various glucopyranosides can be selectively acylated with PPL and *Candida antarctica* lipase (97,98). It is apparent that the high selectivity of lipases toward primary groups can be attributed, at least partly, to higher chemical reactivity of the primary hydroxy group compared to the secondary hydroxyl.

Much more synthetically useful is the ability of lipases and proteases suspended in nonaqueous medium to discriminate between groups that have similar chemical reactivity. This was originally demonstrated with regioselective acylation of sugar derivatives and steroids in organic solvents (98,99). For example, lipases exhibit a remarkable regioselectivity by discriminating among the four secondary hydroxyl groups in C-6 protected glucose (74), galactose,

and mannose. Whereas PPL exclusively acylates the C-2 hydroxyl group, *Chromobacterium viscosum* lipase displays an overwhelming preference for the C-3 hydroxyl group (Table 7). Following this methodology C-2 and C-3 monoesters of glucose have been obtained on a gram scale (98).

Table 7. Regioselective Acylation of Hydroxycompounds

Substrate	Enzyme	Yield, %	Isolated product	Refs.
	PPL	51	6- <i>O</i> -acylglucopyranose	96
	lipase <i>Candida antarctica</i>	85–95	6- <i>O</i> -acylglucopyranoside	97,98
	PPL	51	2,6-di- <i>O</i> -butylglucose	98
	<i>Chromobacterium viscosum</i> lipase	80	3,6-di- <i>O</i> -butylglucose	
	<i>Chromobacterium viscosum</i> lipase	83	3β-monobutyl ester	99
	subtilisin	60	17β-monobutyl ester	99
	subtilisin	45	6'-monobutylmaltose	100
	subtilisin	34	6'-monobutylsalicin	100
	subtilisin	82	1- <i>O</i> -acylcastanospermine	101
	PFL	76–97	1(<i>R</i>),2(<i>S</i>)-monoester	102
	CCL	66–75	1(<i>S</i>),2(<i>R</i>)-monoester	
	lipase PS	64–82	3'-carbonates	103

^a B = thymine, uracil, or adenine.

A number of steroids have been regioselectively acylated in a similar manner (99,104). *Chromobacterium viscosum* lipase esterifies 5α-androstane-3β,17β-diol [571-20-0] (75) with 2,2,2-trifluoroethyl butyrate in acetone with high selectivity. The lipase acylates exclusively the hydroxy group in the 3-position giving the 3β-(monobutyl ester) of (75) in 83% yield. In contrast, *Bacillus subtilis* protease (subtilisin) displays a marked preference for the C-17 hydroxyl. *Candida cylindracea* lipase (CCL) suspended in anhydrous benzene regioselectively acylates the 3α-hydroxyl group of several bile acid derivatives (104).

By taking advantage of the remarkable ability of subtilisin [9014-01-1] to remain catalytically active in anhydrous dimethylformamide, a number of carbohydrates and other sugar-related compounds have been regioselectively acylated with trichloroethyl butyrate (100). In the case of maltose [69-79-4] (76) and salicin [138-52-3] (77), for example, acylation occurs exclusively at the C-6' positions.

Castanospermine [79831-76-8] (78) contains four secondary hydroxyl groups and therefore represents a significant challenge for regioselective

modifications. Subtilisin suspended in pyridine acylates the C-1 hydroxy group with remarkable selectivity resulting in 1-*O*-butyrylcastanospermine in 82% isolated yield (101). The 1-*O*-acyl derivatives can be further acylated at the C-7 hydroxyl with *Chromobacterium viscosum* lipase. Organometallic substrates such as (η⁻1,2-benzene-dimethanol)tricarbonylchromium (79) can also be selectively acylated with lipases (102). Lipase from *Pseudomonas fluorescens* acylates (79) resulting in the formation of 1(*R*)2(*S*)-monoester with 93% *ee*. Conversely, *Candida cylindracea* lipase exhibits opposite selectivity giving the 1(*S*),2(*R*)-enantiomer.

Regioselective acylation of the 3'-hydroxyl group of unprotected nucleosides (80) deserves special attention. This reaction is highly unusual because lipase PS catalyzes acylation of the secondary hydroxyl group in the presence of primary hydroxyl group in a highly selective manner. Treatment of the nucleosides with a slight excess of *O*-[(alkyloxy)carbonyl]-oximes in the presence of the lipase in THF at 60°C yields 3'-carbonates as the only product in 65–82% yield (103). Interestingly, the lipase also exhibits preference for the secondary hydroxyl group of nucleosides in the hydrolytic mode. The hydrolysis of 2'-deoxy-3',5'-dihexanoyl pyrimidine nucleosides proceeds exclusively at the secondary hydroxy group resulting in 2'-deoxy-5'-*O*-acylnucleosides (105).

Resolution of Racemic Amines and Amino Acids. Acylases (*EC 3.5.1.14*) are the most commonly used enzymes for the resolution of amino acids. Porcine kidney acylase (PKA) and the fungal *Aspergillus* acylase (AA) are commercially available, inexpensive, and stable. They have broad substrate specificity and hydrolyze a wide spectrum of natural and unnatural *N*-acyl amino acids, with exceptionally high enantioselectivity in almost all cases. Moreover, their enantioselectivity is exceptionally good with most substrates. A general paper on this subject has been published (106) in which the resolution of over 50 *N*-acyl amino acids and analogues is described. Also reported are the stabilities of the enzymes and the effect of different acyl groups on the rate and selectivity of enzymatic hydrolysis. Some of the substrates that are easily resolved on 10–100 g scale are presented in Figure 4 (106). Lipases are also used for the resolution of *N*-acylated amino acids but the rates and optical purities are usually low (107).

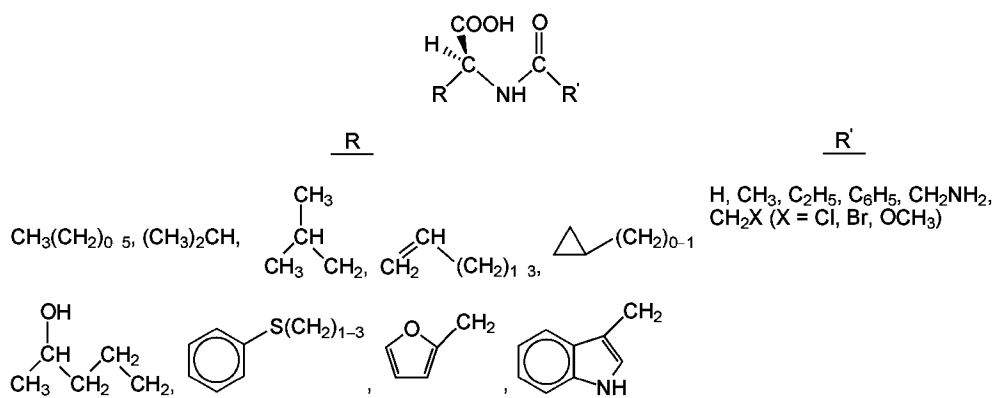
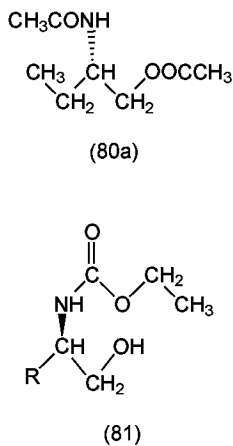


Fig. 4. Examples of enzymatically resolved *N*-acyl amino acids.

Amino alcohols can be resolved by a number of pathways including hydrolysis, esterification, and transesterification. For example, hydrolysis of *N,O*-diacetyl-2-amino-1-butanol with PPL followed by recrystallization results in (80a) with 95% *ee* (108). Hydrolysis of racemic acetates or butyrates of 2-[(alkoxycarbonyl)amino]-1-alkanols with PFL gives (*R*)-alcohol (81) with 95% *ee* (109). (*S*)-(81) can be obtained by transesterification of the racemic (81) with ethyl acetate which also serves as the reaction medium (109).



Unprotected racemic amines can be resolved by enantioselective acylations with activated esters (110,111). This approach is based on the discovery that enantioselectivity of some enzymes strongly depends on the nature of the reaction medium. For example, the enantioselectivity factor (defined as the ratio of the initial rates for (*S*)- and (*R*)-isomers) of subtilisin in the acylation of α-methyl-benzylamine with trifluoroethyl butyrate varies from 0.95 in toluene to 7.7 in 3-methyl-3-pentanol (110). The latter solvent has been used for enantioselective resolutions of a number of racemic amines (110).

Two interesting approaches for resolution of racemic amino acids have been reported (112,113). Both are based on enantioselective ring opening with *in situ* racemization of the nonreactive isomer (Fig. 5). The enantioselective hydrolysis of oxazolidinone (82) by cells of *Pseudomonas testosteroni*, coupled with *Bacillus subtilis*-catalyzed racemization of unreacted substrate, results in the production of L-serine [56-45-1] (84). A more versatile and technically simpler system is based on the enantioselective esterification of azlactones (85) that undergo ring opening with various nucleophiles, including alcohols, and racemization of the unreacted isomer. These two procedures allow, at least in principle, a quantitative conversion of a racemic azlactone to a single isomer of the product, thus overcoming the usual problem of 50% maximum yield of kinetic resolutions.

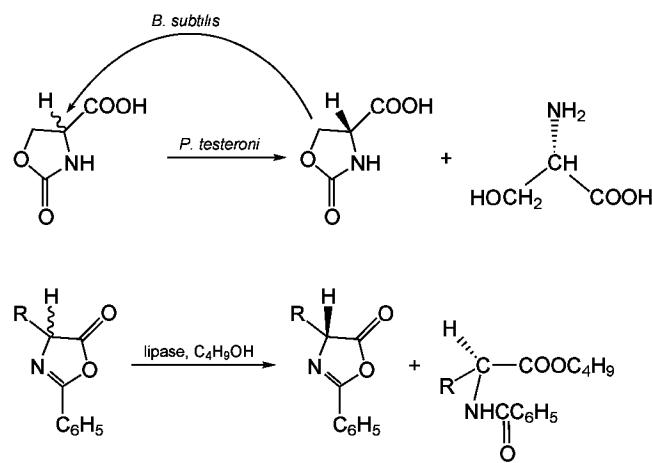
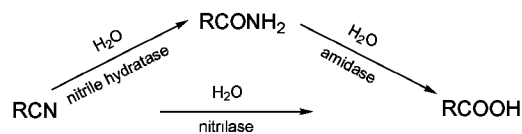


Fig. 5. Enzymatic resolution of amino acids by ring-opening reaction.

Hydrolysis of Nitriles. The chemical hydrolysis of nitriles to acids takes place only under strong acidic or basic conditions and may be accompanied by formation of unwanted and sometimes toxic by-products. Enzymatic hydrolysis of nitriles by nitrile hydratases, nitrilases, and amidases is often advantageous since amides or acids can be produced under very mild conditions and in a stereo- or regioselective manner (114,115). There are two distinct classes of enzymes that hydrolyze nitriles. Nitrilases (*EC 3.5.5.1*) hydrolyze nitriles directly to corresponding acids and ammonia without forming the amide. In fact, amides are not substrates for these enzymes. Nitriles also may be first hydrated by nitrile hydratases to yield amides which are then converted to carboxylic acid with amidases. This is a two-enzyme process, in which enantioselectivity is generally exhibited by the amidase, rather than the hydratase.



The hydrolysis of nitriles can be carried out with either isolated enzymes or immobilized cells. For example, resting cells of *P. chlororaphis* can accumulate up to 400 g L of acrylamide in 8 h, provided acrylonitrile is added gradually to avoid nitrile hydratase inhibition (116). The degree of acrylonitrile conversion to acrylamide is 99% without any formation of acrylic acid. Because of its high efficiency the process has been commercialized and currently is used by Nitto Chemical Industry Co. on a multithousand ton scale. A crude mixture of enzymes isolated from *Rhodococcus* sp. is used for selective hydrolysis of aromatic and aliphatic nitriles and dinitriles (117). Nitrilase accepts a wide range of substrates (Table 8). Even though many of them have low solubility in water, such as (88), the yields are in the range of 90%. Carboxylic esters are not susceptible to the hydrolysis by the enzyme so that only the cyano group of (89) is hydrolyzed. This mode of selectivity is opposite to that observed upon the chemical hydrolysis: at alkaline pH, esters are more labile than nitriles. Dinitriles (90,91) can be hydrolyzed regioselectively resulting in cyanoacids in 71–91% yield. Hydrolysis of (92) proceeds via the formation of racemic amide which is then hydrolyzed to the acid in 95% *ee* (118). Prochiral 3-substituted glutaronitriles (93) are hydrolyzed by *Rhodococcus butanica* in up to 71% yield with excellent selectivity (119).

Table 8. Enzymatic Hydrolysis of Nitriles

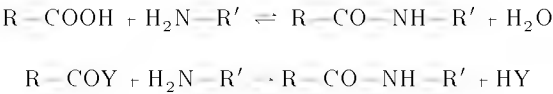
Substrate	Product	Yield, %	<i>ee</i> , %	References
		91		117
		92		117
		71		117
		91		117
		47	95	118
		68–71	90–99	119


$$^a \text{R} = \text{OCH}_2\text{C}_2\text{H}_5, \text{OCOC}_2\text{H}_5.$$

Peptide Synthesis. The literature on the enzymatic synthesis of peptides is enormous (120–124). Here the basic principles that govern peptide synthesis are illustrated and recent trends in this area reviewed.

Despite significant progress in the chemical synthesis of peptides and proteins by both liquid- and solid-phase methodologies, a number of shortcomings still exist. The principal limitation of chemical methods stems from the formation of by-products at each individual condensation step that accumulate during the course of the repeated reactions. Moreover, even the small amount of racemization that often occurs in the presence of highly activated coupling reagents reduces the purity of final product and complicates purification dramatically. In this respect the use of enzymes for amino acid or peptide coupling has a number of advantages. Mild reaction conditions and the excellent stereo- and regioselectivity of enzymes require only minimal protection, precludes racemization, and guarantees the structural fidelity of the product.

There are two basic strategies for enzyme-catalyzed peptide synthesis: equilibrium- and kinetically controlled synthesis. The former is the direct reversal of proteolysis and involves the condensation of an amino component with unactivated carboxyl component. The latter proceeds by the aminolysis of an activated peptide ester.



For the equilibrium-controlled enzyme-catalyzed peptide synthesis the equilibrium position lies far over in the direction of the hydrolysis, and under physiological conditions, the product yield is negligible. The equilibrium position is determined exclusively by thermodynamic factors and like any other catalysts the enzymes only accelerate the attainment of the equilibrium.

In order to obtain a reasonable product yield, a number of approaches have been developed to shift the equilibrium toward condensation. The addition of organic cosolvents, for example, lowers the dielectric constant of the reaction medium, thus reducing the pK difference between α -carboxyl and α -amino groups. Since the predominant contribution to the energetic barrier of peptide condensation is the energy required for proton transfer, this decrease in ΔpK lowers the free energy (ΔG_{syn}) of the reaction (125). ΔG_{syn} can also be minimized by carrying out the reaction at a thermodynamically optimal pH value. It can be shown (121) that pH_{opt} depends on pK of the carboxyl (pK_1) and the amino (pK_2) component and obeys the following equation: $pH_{\text{opt}} = 0.5(pK_1 + pK_2)$. Ionic equilibrium can also be perturbed in favor of peptide synthesis by elevation of the temperature.

The most frequently used technique to shift the equilibrium toward peptide synthesis is based on differences in solubility of starting materials and products. Introduction of suitable apolar protective groups or increase of ionic strength decreases the product solubility to an extent that often allows nearly quantitative conversions. Another solubility-controlled technique is based on introduction of a water-immiscible solvent to give a two-phase system. Products preferentially partition away from the reaction medium thereby shifting the equilibrium toward peptide synthesis.

Although equilibrium-controlled peptide synthesis has been successfully used on a number of occasions, including thermolysin-catalyzed synthesis of aspartame (126) and semisynthesis of insulin (127), the method has a significant drawback; a water-miscible organic cosolvent added to the reaction medium to suppress the ionization of unactivated carboxy components significantly reduces the reaction rate.

Kinetically controlled enzymatic coupling is considered to be more efficient despite the fact that hydrolysis of the growing polypeptide chain may reduce the yield of desired product. A number of interesting approaches have appeared recently that overcome the problem of secondary hydrolysis and increase product yield and purity. They include the use of high concentrations of specific water-miscible organic cosolvents that inhibit amidase activity to a much higher degree than esterase activity (128), the use of proteases and lipases in nonaqueous solvents (129–132), the use of chemically modified proteases with suppressed amidase activity (133,134), and the use of site-specific mutagenesis for design of stable protease mutants that are active at high concentrations of organic cosolvents (135). Several examples of these methodologies are presented in Table 9.

Table 9. Enzyme-Catalyzed Synthesis of Peptides^a

Entry	Carboxy terminal	Amino terminal	Product	Enzyme ^b	Solvent	Yield, %	Refs
1	Boc-Cys(SBzl)-OMe	D Val-OBzl	Boc-Cys(SBzl)-D Val-OBzl	CT	50% DMSO-water	72	128
2	Z-Tyr-OMe	D Arg-OMe	Z-Tyr-D Arg-OMe	CT	50% CH ₃ CN-water	72	128
3	Z-L Phe-OCH ₂ CN	L Leu-NH ₂	Z-L Phe-L Leu-NH ₂	Me-CT	50% DMSO-water	88	133
4	Z-L Phe-C ₆ H ₄ Cl	Gly-Gly-NH ₂	Z-L Phe-Gly-Gly-NH ₂	anhydrosubtilisin	60% DMF-water	90	134
5	Z-Gly-Pro-Phe-Pro-Leu	Leu-NH ₂	Z-Gly-Pro-Phe-Pro-Leu-Leu-NH ₂	thermolysin	<i>tert</i> -amyl alcohol, 9% formamide	73	130
6	Z-Leu-Leu-OMe	Phe-Leu-O(CH ₃) ₃	Z-Leu-Leu-Phe-Leu-O(CH ₃) ₃	subtilisin BNP'	acetonitrile	95	136
7	Bzl-Tyr-OC ₂ H ₅	L Phe-NH ₂	Bzl-Tyr-Phe-NH ₂	CT	1,1,1-tri-chloroethane	95-98	132
8	N-Ac-L Phe-OC ₂ H ₄ Cl	L Leu-NH ₂	N-Ac-L Phe-L Leu-NH ₂	PPL	toluene	83	129
9	N-Ac-L Tyr-OC ₂ H ₄ Cl	L Leu-NH ₂	N-Ac-L Tyr-L Leu-NH ₂	PPL	tetrahydrofuran	76	129
10	N-F-D Ala-OC ₂ H ₄ Cl	L Phe-NH ₂	N-F-D Ala-L Phe-NH ₂	subtilisin	<i>tert</i> -amyl alcohol	82	131

^a Boc - *t*-butoxycarbonyl; Bzl - benzyl; Me - methyl; Z - carbobenzoxy; N-Ac - N-acyl and -F - formyl.
^b CT - α-chymotrypsin; PPL - porcine pancreatic lipase; and Me-CT - chymotrypsin methylated at N^ε of His 57.

Hydrolysis of esters and amides by enzymes that form acyl enzyme intermediates is similar in mechanism but different in rate-limiting steps. Whereas formation of the acyl enzyme intermediate is a rate-limiting step for amide hydrolysis, it is the deacylation step that determines the rate of ester hydrolysis. This difference allows elimination of the undesirable amidase activity that is responsible for secondary hydrolysis without affecting the rate of synthesis. Addition of an appropriate cosolvent such as acetonitrile, DMF, or dioxane can selectively eliminate undesirable amidase activity (128).

The first two entries in Table 9 illustrate the above principle. Both tripeptide Boc-Cys(SBzl)-D-Val-OBzl and dipeptide Z-Tyr-D-Arg-OMe are

enkephalin precursors and can be synthesized in 72% yield, with no secondary hydrolysis observed during the course of the reaction. An alternative way to inhibit amidase activity is based on modifications of residues that are directly involved in catalytic steps, such as chymotrypsin methylated at N⁶² of histidine 57, which is inert toward amide substrates. This modified enzyme has been used as a catalyst for the synthesis of a number of peptides, including Z L Phe L Leu NH₂ in high yield (133). Similarly, the Ser residue of the active site of subtilisin BNP' has been converted to dehydroalanine (134). The resulting anhydrosubtilisin was found to be a useful catalyst for the condensation of peptide fragments. Z L Phe Gly Gly NH₂ was synthesized from the corresponding fragments (Table 9) in 90% yield.

Another useful strategy that helps to eliminate hydrolysis and also shift the thermodynamic equilibrium toward peptide-bond formation is the use of nonaqueous systems (Table 9, entries 5–10). Remarkably, not only proteases such as thermolysin [9073-78-3], subtilisin, and chymotrypsin but also lipases (porcine pancreatic lipase) are capable of catalyzing the formation of peptide bonds under these conditions. The use of lipases may be advantageous in some instances since, in contrast to proteases, they are unable to catalyze the secondary hydrolysis of amide bonds. Another payoff of conducting peptide condensation in nonaqueous media stems from the fact that stereospecificity of enzymes in organic solvents can be significantly altered (137). As a result, peptides containing D-amino acids can be synthesized efficiently under nonaqueous environments (131). An often-cited drawback of the use of organic solvents is that they reduce the enzymatic activity.

Lyases

Aldol Additions. These reactions catalyzed by lyases are perhaps the most synthetically useful enzymatic reactions for carbon–carbon bond formation. Because of the broad synthetic utility of this method, the enzymatic aldol reactions have received considerable attention in recent years and have been extensively covered in a number of books and reviews (10,138–140).

There are two distinct groups of aldolases. Type I aldolases, found in higher plants and animals, require no metal cofactor and catalyze aldol addition via Schiff base formation between the lysine ε-amino group of the enzyme and a carbonyl group of the substrate. Class II aldolases are found primarily in microorganisms and utilize a divalent zinc to activate the electrophilic component of the reaction. The most studied aldolases are fructose-1,6-diphosphate (FDP) enzymes from rabbit muscle, rabbit muscle adolase (RAMA), and a Zn²⁺-containing aldolase from *E. coli*. *In vivo* these enzymes catalyze the reversible reaction of D-glyceraldehyde-3-phosphate [591-57-1] (G-3-P) and dihydroxyacetone phosphate [57-04-5] (DHAP). Although their requirement for G-3-P is strict, the enzymes accept a wide range of aldehydes as electrophiles (Fig. 6). The stereospecificity of the reaction is absolute; the configuration of the vicinal diols at the newly formed C-3–C-4 is always D-threo. A great variety of aldol additions have been carried out, allowing the synthesis of numerous nitrogen-containing sugars, deoxysugars, and fluorosugars, based, among others, on D-threose, L-threose, D-ribose, L-arabinose, D-xylose, D-glucose-6-P, D-galactose-6-P, and D-mannose-6-P, where P = phosphate. Some other examples of (94) are (a–j).

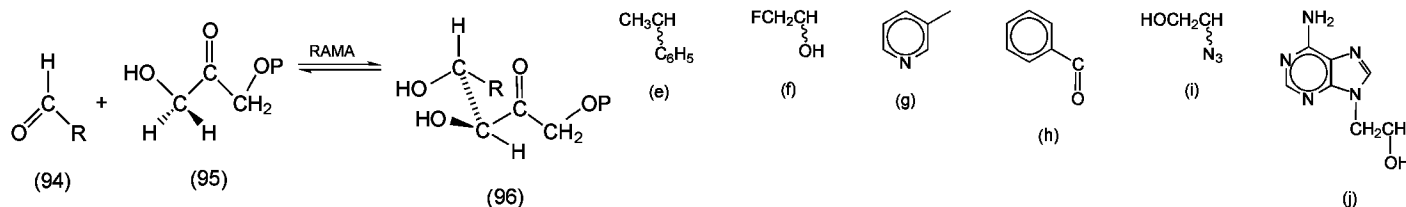
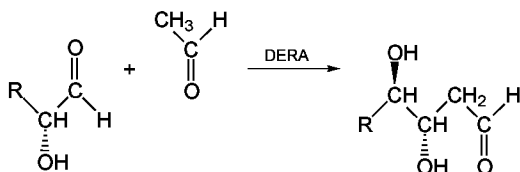
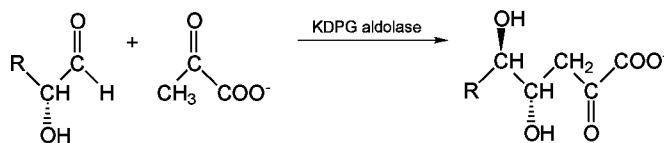


Fig. 6. FDP-aldolase-catalyzed addition of electrophiles (94) with DHAP (139–146). Representative R groups in (94) are given as (a–j): (a) methyl, CH₃; (b) azidomethyl, N₃CH₂; (c) benzyloxymethyl, C₆H₅CH₂OCH₂; (d) carboxaldehyde, CHO.

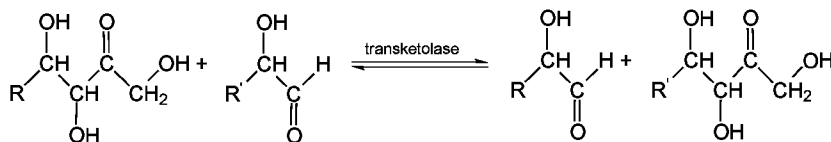
Deoxyribose-5-phosphate aldolase (DERA) and 2-keto-3-deoxy-6-phosphogluconate aldolase (KDPG) are also useful for asymmetric aldol additions. DERA, the only aldolase that utilizes aldehydes as the nucleophile, is capable of catalyzing the synthesis of a wide range of products. In addition to its natural substrate, acetaldehyde, it accepts a number of analogues. When acetaldehyde [75-07-0], acetone [67-64-1], or fluoroacetone are used as substrates, a new stereogenic center with 3(*S*)-configuration is formed. The reaction with propanal [123-38-6] results in two new stereogenic centers with 2(*R*)- and 3(*S*)-configurations (147).



KDPG is a member of a yet unexplored group of aldolases that utilize pyruvate or phosphoenol pyruvate as the nucleophile in the aldol addition. They are quite tolerant of different electrophilic components and accept a large number of unnatural aldehydes (148). The reaction itself, however, is quite specific, generating a new stereogenic center at the C-4 position.



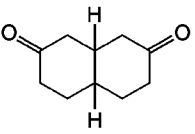
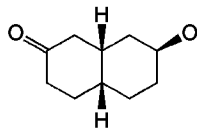
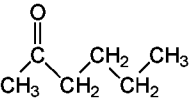
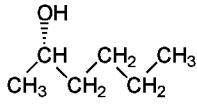
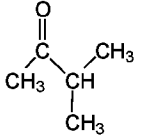
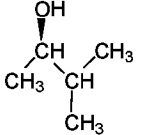
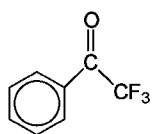
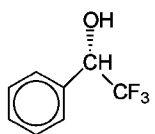
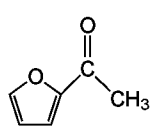
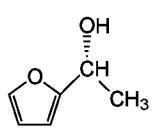
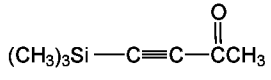
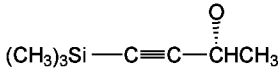
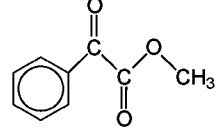
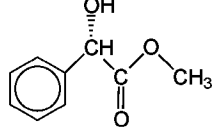
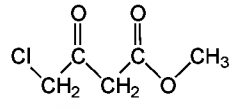
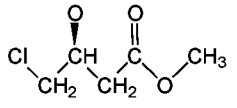
Transketolases (TK) are widely used in organic synthesis to extend the chain of an acceptor aldose by two carbon units.



The TK-catalyzed reaction requires the presence of thiamine pyrophosphate and Mg²⁺ as cofactors. Although the substrate specificity of the enzyme has not been thoroughly investigated, it has been shown that the enzyme accepts a wide variety of 2-hydroxyaldehydes including D-glyceraldehyde 3-phosphate [591-57-1], D-glyceraldehyde [453-17-8], D-ribose 5-phosphate [4300-28-1], D-erythrose 4-phosphate [585-18-2], and D-erythrose [583-50-6] (139,149–151). Most substrates, with the exception of hydroxypyruvate, have a threo configuration of hydroxyl groups at the C-3 and C-4 positions (139). The new stereocenter formed in TK-catalyzed addition is formed in the threo configuration with high diastereo-selectivity (151). Using TK-catalyzed condensations of hydroxypyruvic acid with a number of aldehydes a practical preparative synthesis of L-idose [5934-56-5], L-gulose [6027-89-0], 2-deoxy-L-xylohexose, and

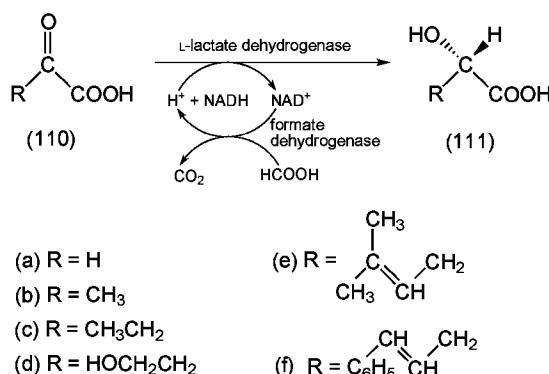
transfer pro-(R)-hydride to the *re* face of the carbonyl group to give (*S*)-alcohols, a process which can be described by Prelog's rule (171). In contrast alcohol dehydrogenase from *L. kefir* (LKADH) and *Pseudomonas* sp. (PEDADH) exhibit anti-Prelog specificity and transfer pro-(R)-hydride to the *si* face of carbonyl compounds to yield (R)-alcohols (169,170). Both enzymes exhibit broad substrate specificity and high enantioselectivity for the synthesis of chiral aromatic (105,108), cyclic (106), aliphatic (109), and trimethylsilyl-protected terminal alkyne (107) ketones.

Table 11. Alcohol Dehydrogenase-Catalyzed Reductions on Ketones

Substrate	Product	Enzyme	Yield, %	ee, %	Refs.
		HLAD	89	>98	167
		TBADH	50 ^a	96	168
		TBADH	50 ^a	86	168
		LKADH	71	>99	169
		LKADH	65	65	169
		LKADH	25	94	169
		PSADH	79	98	170
		PSADH	76	98	170

^a Stopped at 50% conversion.

Hydroxy and Amino Acids. Reduction of 2-oxoacids with NADH, catalyzed by lactate dehydrogenase [9001-60-9], is an established method for the preparation of homochiral α -hydroxy acids of both S and R configurations (172–174). The enzyme is found in all higher organisms and can be easily isolated from a variety of mammalian (172) and bacterial (173,10) sources. The reduction of (110a–d), where (a) glyoxylic acid [298-12-4] (b) pyruvic acid [127-17-3], (c) α -ketobutyric acid [600-18-0], and (d) 4-hydroxy-2-oxobutanoic acid [22136-38-5]), proceeds with excellent yield (97–99%) and selectively (*ee* > 99%). Although reduction of β,γ -unsaturated substrates is relatively slow, it provides synthetically useful functionalized allylic alcohols with high optical purity.

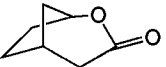
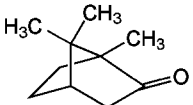
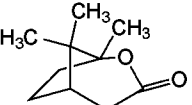
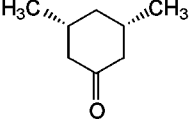
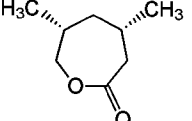
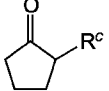
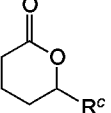


There are several amino acid dehydrogenases that catalyze reductive amination of oxo acids to L-amino acids in nearly quantitative yield (161). The enzymes accept a variety of 2-keto acids and are useful for synthesis of unnatural amino acids. For example, leucine dehydrogenase was used for the synthesis of 2-amino-3,3-dimethylbutanoic acid (175) and phenylalanine dehydrogenase was used for synthesis of L-2-amino-4-phenylbutanoic acid (176), a precursor of angiotensin-converting enzyme inhibitors (qv). In most cases regeneration of NADH was accomplished with formate–formate dehydrogenase or glucose–glucose dehydrogenase systems.

Baeyer-Villiger Oxidations. The Baeyer-Villiger reaction is a useful transformation in organic synthesis providing a mild technique for converting ketones into esters or lactones. The biological equivalent of this transformation is a common reaction that occurs during degradation of steroids, aldehydes, and cyclic ketones. The enzyme that catalyzes the insertion of an oxygen atom is classified as a monooxygenase and uses flavin to facilitate the passage of electrons (177). A number of these enzymes have been purified to homogeneity and characterized.

Cyclohexanone oxygenase from *Acinetobacteria* converts a variety of alicyclic ketones into lactones in a regio- and enantioselective manner (Table 12). The reaction can be carried out by a whole-cell process (181) or with the isolated enzyme (178). For example, 2-norbomanone [497-38-1] (112) is converted to the corresponding lactone in 81% yield (178). The enzyme, however, is not selective: both enantiomers react equally well. The oxidation rate of camphor [76-22-2] (113), is about one-third that of (112); nevertheless, given sufficient amount of time, the product yield reaches 89%. Substituted cyclohexanones (114) and cyclopentanones (115) are converted into the corresponding lactones in moderate to good yield and selectivity (179–181).

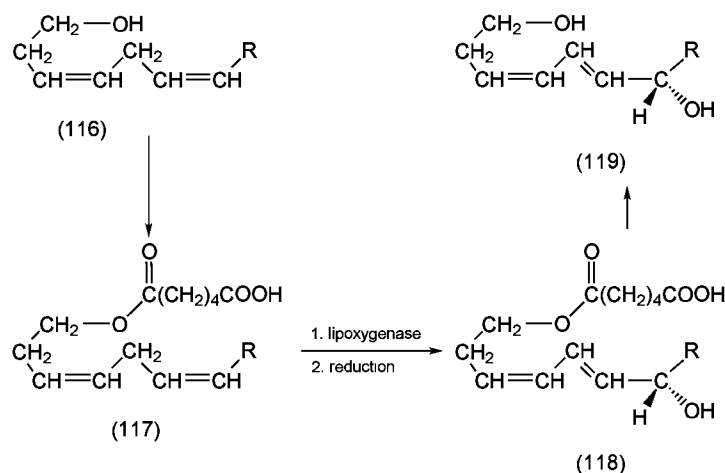
Table 12. Baeyer-Villiger Oxidation of Ketones

	Substrate	Product	Yield, %	References
(112)			81	178
(113)			89	178
(114) ^a			73–82	179,180
(15) ^b			20–65 25–95% ^c	181

^a *cis*-3,5-Dimethylcyclohexanone [7214-52-0].^b 2-Heptylcyclopentanone [137-03-1], 2-nonylcyclopentanone [40566-27-6]. $^c \text{R} = \text{C}_7\text{H}_{15}, \text{C}_9\text{H}_{19}, \text{and } \text{C}_{11}\text{H}_{23}$

Lipoxygenase-Catalyzed Oxidations. Lipoxygenase-1 catalyzes the incorporation of dioxygen into polyunsaturated fatty acids possessing a 1(Z),4(Z)-pentadienyl moiety to yield (E),(Z)-conjugated hydroperoxides. A highly active preparation of the enzyme from soybean is commercially available in purified form. From a practical standpoint it is important to mention that the substrate does not need to be in solution to undergo the oxidation. Indeed, the treatment of 28 g L of linoleic acid [60-33-3] with 2 mg of the enzyme results in (13S)-hydroperoxide of linoleic acid in 80% yield and 95% ee (182).

Another application of lipoxygenase (183) is the hydroxylation of unsaturated alcohols (116) via their adipate monoesters (117) where R is $(\text{CH}_2)_4\text{CH}_3$, $\text{CH}_3\text{CH}(\text{CH}_3)_2$, $\text{CH}_3\text{C}_6\text{H}_5$, $(\text{CH}_2)_3\text{C}_6\text{H}_5$, or $(\text{CH}_2)_3\text{C}(\text{O})\text{CH}_3$.



Adipoyl moiety is first attached in order to provide an appropriately spaced proximal carbonyl group. The esters are oxidized enzymatically and then deacylated. The procedure results in the synthesis of diols (119) with excellent enantiomeric purity (*ee* 96–98° o) in 72–92° o yield.

Summary

The area of biotransformations has experienced a phenomenal growth in recent years as judged by the number of publications. In addition to the fermentation technology that has been utilized for years for the production of molecules varying from ethanol and vitamins to amino acids and steroids, the practical use of isolated enzymes as catalysts in organic synthesis has started to achieve great prominence. This phenomenon has been brought about by a number of factors. In response to an ever-growing demand for new biocatalysts, more enzymes have become available in recent years. Such primary suppliers as Amano, Genencor, Novo, Meito Sangyo, Toyo Jozo, Rhône-Poulenc, Serva, Nagase, Tanabe, Boehringer-Mannheim, and others now trade dozens of enzymes, many in ton quantities. Moreover, advances in molecular biology and genetic engineering have resulted in new enzymes with altered specificity and increased stability. For example, highly stable subtilisin mutants with an increased propensity for peptide condensation (184) have been constructed and successfully used for peptide synthesis (135). A modified lactate dehydrogenase that is specific for new substrates and lacks allosteric regulation (185) has been used for enantioselective reductions (173).

Impressive developments in the area of immunology have culminated in the development of catalytic antibodies (186,187). These synthetic biocatalysts that have the potential to catalyze virtually any type of reaction with unsurpassed selectivity have great promise in the future.

Perhaps the biggest impact on the practical utilization of enzymes has been the development of nonaqueous enzymology (11,16,33,35). The use of enzymes in nonaqueous media greatly expands the scope of suitable transformations, simplifies their use, and enhances stability. It also provides an easy means of regulation of the substrate specificity and regio- and enantioselectivity of enzymes by changing the reaction medium.

It is apparent that the use of enzymatic catalysis continues to grow. Greater availability of enzymes, development of new methodologies for their utilization, investigation of enzymatic behavior in nonconventional environments, and the design and synthesis of new biocatalysts with altered selectivity and increased stability are essential for the successful development of this field. As more is learned about selectivity of enzymes toward unnatural substrates, the choice of an enzyme for a particular transformation will become easier to predict. It should simplify a search for an appropriate catalyst and help to establish biocatalytic procedures as a useful supplement to classical organic synthesis.

BIBLIOGRAPHY

1. L. G. Copping, R. Martin, J. A. Pickett, C. Bucke, and A. W. Bunch, eds., *Opportunities in Biotransformations*, Elsevier, London, 1990.
2. H. G. Davis, R. H. Green, D. R. Kelly, and S. M. Roberts, eds., *Biotransformations in Preparative Organic Chemistry*, Academic Press Ltd. London, 1989.
3. E. Santaniello, P. Ferraboschi, P. Grisenti, and A. Manzocchi, *Chem. Rev.* **92**, 1071 (1992).
4. H. L. Holland, *Organic Synthesis with Oxidative Enzymes*, VCH Publishers, Inc., New York, 1992.
5. L. Zhu and M. C. Tedford, *Tetrahedron* **46**, 6587 (1990).
6. S. Servi, *Synthesis* **1**, 1 (1990).
7. C. J. Sih and S.-H. Wu, *Top. Stereochem.* **19**, 63 (1989).
8. D. A. Abramowicz, ed. *Biocatalysis*, Van Nostrand Reinhold Co., Inc. New York, 1990.
9. J. S. Dordick, ed., *Biocatalysis For Industry*, Plenum Publishing Corp., New York, 1991.
10. M. Bednarski and E. S. Simon, eds., *Enzymes in Carbohydrate Synthesis*, American Chemical Society, Washington, D.C., 1991.
11. A. M. Klibanov, *Accounts Chem. Res.* **23**, 114 (1990).
12. C.-H. Wong, *Science* **244**, 1145 (1989).
13. W. Boland, C. Frossl, and M. Lorenz, *Synthesis*, 1049 (1991).
14. Y. Ichikawa, G. C. Look, and C.-H. Wong, *Anal. Biochem.* **202**, 215 (1992).
15. C.-S. Chen, Y. Fujimoto, G. Girdaukas, and C. J. Sih, *J. Am. Chem. Soc.* **104**, 7294 (1982).
16. C.-H. Chen and C. J. Sih, *Angew. Chem. Int. Ed. Engl.* **28**, 695 (1989).
17. E. Santaniello, M. Chiari, P. Ferraboschi, and S. J. Trave, *J. Org. Chem.* **53**, 1567 (1988).
18. R. Andruszkiewicz, A. G. M. Barrett, and R. B. Silverman, *Synth. Commun.* **20**, 159 (1990).
19. R. Roy and A. W. Rey, *Tetrahedron Lett.* **28**, 4935 (1987).
20. T. Kitazume, N. Okamura, T. Ikeya, and T. J. Yamazaki, *J. Fluorine Chem.* **39**, 107 (1988).
21. A. Fadel, J.-L. Canet, and J. Salaun, *Synlett*, 60 (1991).
22. F. Bjorkling and co-workers, *Tetrahedron* **41**, 1347 (1985).
23. H. J. Bestmann and U. C. Philipp, *Angew. Chem. Int. Ed. Engl.* **30**, 86 (1991).
24. D. L. Hughes and co-workers, *J. Org. Chem.* **55**, 6252 (1990).
25. G. Sabbioni and J. B. Jones, *J. Org. Chem.* **52**, 4565 (1987).
26. M. P. Schneider, N. Engel, P. Honicke, G. Heinemann, and H. Gorisch, *Angew. Chem. Int. Ed. Engl.* **23**, 67 (1984).
27. R. Bloch, E. Guibe-Jample, and C. Girard, *Tetrahedron Lett.* **26**, 4087 (1985).
28. S. Niwayama, S. Kobayashi, and M. Ohno, *Tetrahedron Lett.* **29**, 6313 (1988).
29. H.-J. Gais and co-workers, *J. Org. Chem.* **54**, 5115 (1989).
30. W. E. Ladner and G. M. Whitesides, *J. Am. Chem. Soc.* **106**, 7250 (1984).
31. E. J. Toone, M. J. Werth, and J. B. Jones, *J. Am. Chem. Soc.* **112**, 4946 (1990).
32. P. G. Hultin, F.-J. Mueseler, and J. B. Jones, *J. Org. Chem.* **56**, 5375 (1991).
33. A. Zaks in Ref. 9, p. 161.
34. K. Faber and S. Riva, *Synthesis*, 895 (1992).

35. J. S. Dordick, *Enzyme Microb. Technol.* **11**, 194 (1989).

36. A. Zaks and A. M. Klibanov, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3192 (1985).

37. A. Gutman, M. Shapira, and A. Boltanski, *J. Org. Chem.* **67**, 1063 (1992).

38. Y. Yamamoto, K. Yamamoto, T. Nishioka, and J. Oda, *Agric. Biol. Chem.* **52**, 3087 (1988).

39. G. Guanti and co-workers, *Tetrahedron* **46**, 7081 (1990).

40. K. Mori and Y. Takahashi, *Liebigs Ann. Chem.*, 1057 (1991).

41. G. Guanti, L. Banfi, and E. Narisano, *Tetrahedron Lett.* **31**, 6421 (1990).

42. B. Wirz, R. Schmid, and J. Foricher, *Tetrahedron: Asymmetry* **3**, 137 (1992).

43. U. Ader, D. Breitgoff, P. Klein, K. E. Laumen, and M. P. Schneider, *Tetrahedron Lett.* **30**, 1793 (1989).

44. H. Estermann and co-workers, *Tetrahedron Lett.* **31**, 445 (1990).

45. H. Suemune, K. Matsuno, M. Uchida, and K. Sakai, *Tetrahedron: Asymmetry* **3**, 297 (1992).

46. T. Sugai and K. Mori, *Synthesis*, 19 (1988).

47. B. Danieli, G. Lesma, M. Mauro, G. Palmisano, and D. Passarella, *Tetrahedron: Asymmetry* **1**, 793 (1990).

48. J. Ehler and D. Seebach, *Liebigs Ann. Chem.*, 379 (1990).

49. Y.-F. Wang and C.-H. Wong, *J. Org. Chem.* **53**, 3127 (1988).

50. E. Santaniello, P. Ferraboschi, and P. Grisenti, *Tetrahedron Lett.* **31**, 5657 (1990).

51. F. Theil, H. Schick, G. Winter, and G. Reck, *Tetrahedron* **47**, 7569 (1991).

52. K. J. Harris, Q.-M. Gu, Y.-E. Shih, G. Girdaukas, and C. J. Sih, *Tetrahedron Lett.* **32**, 3941 (1991).

53. F. Theil, H. Schick, M. A. Lapitskaya, and K. K. Pivnitsky, *Liebigs. Ann. Chem.*, 195 (1991).

54. M. Pottie, J. Van der Eycken, and M. Vandewalle, *Tetrahedron: Asymmetry* **2**, 329 (1991).

55. H.-J. Gais, H. Hemmerle, and S. Kossek, *Synthesis*, 169 (1992).

56. D. L. Delinck and A. L. Margolin, *Tetrahedron Lett.* **31**, 6797 (1990).

57. P. Kalaritis, R. W. Regenye, J. J. Partridge, and D. L. Coffen, *J. Org. Chem.* **55**, 812 (1990).

58. R. Chenevert and L. D'Astous, *Can. J. Chem.* **66**, 1219 (1988).

59. F. J. Urban, R. Breitenbach, and L. A. Vincent, *J. Org. Chem.* **55**, 3670 (1990).

60. C. Feichter, K. Faber, and H. J. Griengel, *J. Chem. Soc., Perkin Trans.* **1**, 653 (1991).

61. C. J. Sih, Q.-M. Gu, G. Fulling, S.-H. Wu, and D. R. Reddy, *Developments in Industrial Microbiology* **29**, 221 (1988).

62. K. Burgess and I. Henderson, *Tetrahedron Lett.* **30**, 3633 (1989).

63. R. L. Pederson, K. K.-C. Liu, J. F. Rutan, L. Chen, and C.-H. Wong, *J. Org. Chem.* **55**, 4897 (1990).

64. C. Yee, T. A. Blythe, T. J. McNabb, and A. E. Walts, *J. Org. Chem.* **57**, 3525 (1992).

65. K. Burgess, I. Henderson, and K.-K. Ho, *J. Org. Chem.* **57**, 1290 (1992).

66. S. Liang, and L. A. Paquette, *Tetrahedron: Asymmetry* **1**, 445 (1990).

67. E. Foelsche, A. Hickel, H. Honig, and P. J. Seuffer-Wasserthal, *J. Org. Chem.* **55**, 1749 (1990).

68. H. S. Bevinakatti and A. A. Banerji, *J. Org. Chem.* **56**, 5372 (1991).

69. D. Bianchi, W. Cabri, P. Cesti, F. Francalanci, and F. Rama, *Tetrahedron Lett.* **29**, 2455 (1988).

70. H. Akita, Y. Enoki, H. Yamada, and T. Oishi, *Chem. Pharm. Bull.* **37**, 2876 (1989).

71. H. Ebüke, K. Maruyama, and K. Achiwa, *Tetrahedron: Asymmetry* **3**, 1153 (1992).

72. H. Honig and P. J. Seuffer-Wasserthal, *Synthesis*, 1137 (1990).

73. T. Itoh, Y. Takagi, and S. Nishiyama, *J. Org. Chem.* **56**, 1521 (1991).

74. D. Bianchi, P. Cesti, and E. Battistel, *J. Org. Chem.* **53**, 1988 (5531).

75. S.-H. Hsu, S.-S. Wu, Y.-F. Wang, and C.-H. Wong, *Tetrahedron Lett.* **31**, 6403 (1990).

76. B. Morgan, A. C. Oehlschlager, and T. M. Stokes, *J. Org. Chem.* **57**, 3231 (1992).

77. K. Burgess and L. D. Jennings, *J. Am. Chem. Soc.* **113**, 6129 (1991).

78. M.-J. Kim and Y. K. Choi, *J. Org. Chem.* **57**, 1605 (1992).

79. A. J. M. Janssen, A. J. H. Klunder, and B. Zwanenburg, *Tetrahedron* **47**, 5513 (1991).

80. A. J. M. Janssen, A. J. H. Klunder, and B. Zwanenburg, *Tetrahedron Lett.* **31**, 7219 (1990).

81. R. Seemayer and M. P. Schneider, *Recl. Trav. Chim. Pays-Bas* **110**, 171 (1991).

82. R. J. Cregge, E. R. Wagner, J. Freedman, and A. L. Margolin, *J. Org. Chem.* **55**, 4237 (1990).

83. K. A. Babiak and co-workers, *J. Org. Chem.* **55**, 3377 (1990).

84. K. Matsumoto, S. Tsutsumi, T. Ihori, and H. Ohta, *J. Am. Chem. Soc.* **112**, 9614 (1990).

85. K. Matsumoto and H. Ohta, *Chem. Lett.*, 1589 (1989).

86. K. Matsumoto, N. Suzuki, and H. Ohta, *Tetrahedron Lett.* **31**, 7159 (1990); **31**, 7163 (1990).

87. A. L. Gutman, K. Zuobi, and A. Boltansky, *Tetrahedron Lett.* **28**, 3861 (1987).

88. A. L. Gutman, K. Zuobi, and T. Bravdo, *J. Org. Chem.* **55**, 3546 (1990).

89. C. Bonini, P. Pucci, and L. Viggiani, *J. Org. Chem.* **56**, 4050 (1991).

90. Z. W. Guo, T. K. Ngooi, A. Scilimati, G. Fulling, and C. J. Sih, *Tetrahedron Lett.* **29**, 4927 (1988).

91. G. Zhi-Wei and C. J. Sih, *J. Am. Chem. Soc.* **110**, 1999 (1988).

92. A. L. Margolin, J.-Y. Crene, and A. M. Klibanov, *Tetrahedron Lett.* **28**, 1607 (1987).

93. P. Cesti, A. Zaks, and A. M. Klibanov, *Appl. Biochem. Biotechnol.* **11**, 401 (1985).

94. G. Ottolina, G. Carrea, and S. Riva, *J. Org. Chem.* **55**, 2366 (1990).

95. S. Ramaswamy, B. Morgan, and A. C. Oehlschlager, *Tetrahedron Lett.* **31**, 3405 (1990).

96. M. Therisod and A. M. Klibanov, *J. Am. Chem. Soc.* **108**, 5638 (1986).

97. F. Bjorkling, S. E. Godtfredsen, and O. Kirk, *J. Chem. Soc., Chem. Commun.*, 934 (1989).

98. M. Therisod and A. M. Klibanov, *J. Am. Chem. Soc.* **109**, 3977 (1987).

99. S. Riva and A. M. Klibanov, *J. Am. Chem. Soc.* **110**, 3291 (1988).

100. S. Riva, J. Chopineau, A. P. G. Kieboom, and A. M. Klibanov, *J. Am. Chem. Soc.* **110**, 584 (1988).

101. A. L. Margolin, D. L. Delinck, and M. R. Whalon, *J. Am. Chem. Soc.* **112**, 2849 (1990).

102. Y. Yamazaki and K. Hosono, *Agric. Biol. Chem.* **54**, 3357 (1990).

103. F. Moris and V. Gotor, *J. Org. Chem.* **57**, 2490 (1992).

104. S. Riva, R. Bovara, G. Ottolina, F. Secundo, and G. Carrea, *J. Org. Chem.* **54**, 3161 (1989).

105. A. Uemura, K. Nozaki, J. Yamashita, and M. Yasumoto, *Tetrahedron Lett.* **30**, 3819 (1989).

106. H. K. Chenault, J. Dahmer, and G. Whitesides, *J. Am. Chem. Soc.* **111**, 6354 (1989).

107. T. Miazawa, T. Takitani, S. Ueji, T. Yamada, and S. Kuwata, *J. Chem. Soc., Chem. Commun.*, 1214 (1988).

108. H. S. Bevinakatti and R. V. Newadkar, *Tetrahedron: Asymmetry* **1**, 583 (1990).

109. F. Francalanci, P. Cesti, W. Cabri, D. Bianchi, T. Martinengo, and M. Foa, *J. Org. Chem.* **52**, 5079 (1987).

110. H. Kitaguchi, P. A. Fitzpatrick, J. E. Huber, and A. M. Klibanov, *J. Am. Chem. Soc.* **111**, 3094 (1989).

111. A. L. Gutman, E. Meyer, E. Kalerin, F. Polyak, and J. Sterling, *Biotechnol. Bioeng.* **40**, 760 (1992).

112. K. Yokozeki, E. Majima, K. Izawa, and K. Kubita, *Agric. Biol. Chem.* **51**, 729 (1987).

113. H. S. Bevinakatti, R. V. Newadkar, and A. A. Banerji, *J. Chem. Soc., Chem. Commun.*, 1091 (1990).
114. L. A. Thompson, C. J. Knowels, E. A. Linton, and J. M. Wyatt, *Chem. Brit.* **24**, 900 (1988).
115. T. Nagasawa and H. Yamada, *Pure Appl. Chem.* **62**, 1441 (1990).
116. T. Nagasawa and H. Yamada, *TIBTECH* **7**, 153 (1989).
117. M. A. Cohen, J. Sawden, and N. J. Turner, *Tetrahedron. Lett.* **31**, 7223 (1990).
118. D. Bianchi, A. Bosetti, P. Cesti, G. Franzosi, and S. Spezia, *Biotechnol. Lett.* **13**, 241 (1991).
119. H. Takeya, N. Sakai, A. Sano, M. Yokoyama, T. Sugai, and H. Ohta, *Chem. Lett.*, 1823 (1991).
120. H.-D. Jakubke, P. Kuhl, and A. Konnecke, *Angew. Chem. Int. Ed. Engl.* **24**, 85 (1985).
121. W. Kullman, *Enzymatic Peptide Synthesis*, CRC Press, Boca Raton, Fla., 1987.
122. H.-D. Jakubke in S. Udenfriend and J. Meienhofer, eds., *The Peptides: Analysis, Synthesis, Biology*, Vol. 9., Academic Press, Inc., New York, 1987, p. 103.
123. V. Kasche, in R. J. Beynon and J. S. Bond, eds., *Proteolytic Enzymes: A Practical Approach*, IRL, Oxford, 1989, p. 125.
124. V. Schellenberger and H.-D. Jakubke, *Angew. Chem. Int. Ed. Engl.* **30**, 1437 (1991).
125. G. A. Homandberg, J. A. Mattis, and M. Laskovsky, Jr., *Biochemistry* **17**, 5220 (1978).
126. Y. Isowa and co-workers, *Tetrahedron Lett.* **28**, 2611 (1970).
127. K. Inouye and co-workers, *J. Am. Chem. Soc.* **101**, 751 (1979).
128. C. F. Barbas, III, J. R. Matos, J. B. West, and C.-H. Wong, *J. Am. Chem. Soc.* **110**, 5162 (1988).
129. A. L. Margolin and A. M. Klibanov, *J. Am. Chem. Soc.* **109**, 3802 (1987).
130. H. Kitaguchi and A. M. Klibanov, *J. Am. Chem. Soc.* **111**, 9272 (1989).
131. A. L. Margolin, D.-F. Tai, and A. M. Klibanov, *J. Am. Chem. Soc.* **109**, 7885 (1987).
132. H. Gaertner, T. Watanabe, J. V. Sinisterra, and A. Puigserver, *J. Org. Chem.* **56**, 3149 (1991).
133. J. B. West, J. Scholten, N. J. Stolowich, J. L. Hogg, A. I. Scott, and C.-H. Wong, *J. Am. Chem. Soc.* **110**, 3709 (1988).
134. K. Tanizawa, A. Sugimura, and Y. Kanaoka, *FEBS Lett.* **296**, 163 (1992).
135. C.-H. Wong and co-workers, *J. Am. Chem. Soc.* **112**, 945 (1990).
136. P. Wang, and co-workers, *J. Am. Chem. Soc.* **114**, 378 (1992).
137. P. A. Fitzpatrick, and A. M. Klibanov, *J. Am. Chem. Soc.* **113**, 3166 (1991).
138. D. G. Drueckhammer and co-workers, *Synthesis*, 499 (1991).
139. E. J. Toone, E. S. Simon, M. D. Bednarski, and G. M. Whitesides, *Tetrahedron* **45**, 5365 (1989).
140. C.-H. Wong and co-workers in D. Horton, L. D. Hawkins, G. J. McGarvey, eds., *Trends in Synthetic Carbohydrate Chemistry*, ACS Symposium Series, No. 386, ACS, Washington, D.C., p. 317.
141. M. D. Bednarski and co-workers, *J. Am. Chem. Soc.* **111**, 627 (1989).
142. J. R. Durrwachter, D. G. Drueckhammer, K. Nozaki, H. M. Sweers, and C.-H. Wong, *J. Am. Chem. Soc.* **108**, 7812 (1986).
143. M. D. Bednarski, H. J. Waldman, and G. M. Whitesides, *Tetrahedron Lett.* **27**, 5807 (1986).
144. K. K.-C. Liu and C.-H. Wong, *J. Org. Chem.* **57**, 4789 (1992).
145. R. R. Hung, J. A. Straub, and G. M. Whitesides, *J. Org. Chem.* **56**, 3849 (1991).
146. W.-D. Fessner, J. Badia, O. Eyrisch, A. Schneider, and G. Sinerius, *Tetrahedron Lett.* **33**, 5231 (1992).
147. L. Chen, D. P. Dumas, and C.-H. Wong, *J. Am. Chem. Soc.* **114**, 741 (1992).
148. S. T. Allen, G. R. Heintzelman, and E. J. Toone, *J. Org. Chem.* **57**, 426 (1992).
149. J. Bolte, C. Demuyneck, and H. Samaki, *Tetrahedron Lett.* **28**, 5525 (1987).
150. C. Demuyneck, J. Bolte, L. Hequet, and V. Dalmás, *Tetrahedron Lett.* **32**, 5085 (1991).
151. Y. Kobori, D. C. Myles, and G. M. Whitesides, *J. Org. Chem.* **57**, 5899 (1992).
152. F. Effenberger, T. Ziegler, and S. Forster, *Angew. Chem. Int. Ed. Engl.* **26**, 458 (1987).
153. F. Effenberger, B. Horsch, S. Forster, and T. Ziegler, *Tetrahedron Lett.* **31**, 1249 (1990).
154. F. Effenberger, B. Horsch, F. Weingart, T. Ziegler, and S. Kuhner, *Tetrahedron Lett.* **32**, 2605 (1991).
155. F. Effenberger and U. Stelzer, *Angew. Chem. Int. Ed. Engl.* **30**, 873 (1991).
156. V. I. Ognyanov, V. K. Datcheva, and K. Kyler, *J. Am. Chem. Soc.* **113**, 6992 (1991).
157. T. H. Huuhtanen and L. T. Kanerva, *Tetrahedron: Asymmetry* **3**, 1223 (1992).
158. J. Brussee, E. C. Roos, and A. Van der Gen, *Tetrahedron Lett.* **29**, 4485 (1988).
159. U. Niedermeyer and M.-R. Kula, *Angew. Chem. Int. Ed. Engl.* **29**, 386 (1990).
160. H. K. Chenault and G. M. Whitesides, *Appl. Biochem. Biotechnol.* **14**, 147 (1987).
161. W. Hummel and M.-R. Kula, *Eur. J. Biochem.* **184**, 1, (1989).
162. A. Honorat-Pascal, F. Monot, and D. Ballerini, *Appl. Microbiol. Biotechnol.* **34**, 236 (1990).
163. R. Csuk and B. I. Glanzer, *Chem. Rev.* **91**, 49 (1991).
164. B. J. Jones, and I. Y. Jacobov, *Organic Synthesis*, Vol. 63, John Wiley & Sons, Inc., New York, 1985, p. 10.
165. I. J. Jakovac, H. B. Goodbrand, K. P. Lok, and J. B. Jones, *J. Am. Chem. Soc.* **104**, 1982 (4659).
166. C.-H. Wong and J. R. Matos, *J. Org. Chem.* **50**, 1992 (1985).
167. D. R. Dodds and J. B. Jones, *J. Am. Chem. Soc.* **110**, 577 (1988).
168. E. Keinan, E. K. Hafeli, K. K. Seth, and R. Lamed, *J. Am. Chem. Soc.* **108**, 162 (1986).
169. C. W. Bradshaw, W. Hummel, and C.-H. Wong, *J. Org. Chem.* **57**, 1532 (1992).
170. C. W. Bradshaw, H. Fu, G.-J. Shen, and C.-H. Wong, *J. Org. Chem.* **57**, 1526 (1992).
171. V. Prelog, *Pure Appl. Chem.* **9**, 119 (1964).
172. M.-J. Kim and G. M. Whitesides, *J. Am. Chem. Soc.* **110**, 2959 (1988).
173. G. Casy, T. Lee, and H. Lovell, *Tetrahedron Lett.* **33**, 817 (1992).
174. G. Casy, and co-workers, *J. Chem. Soc., Chem. Commun.*, 924 (1992).
175. C. Wandrey and B. Bossov, *Biotechnol. Bioeng.* **3**, 8 (1986).
176. C. W. Bradshaw, C.-H. Wong, W. Hummel, and M.-R. Kula, *Bioorg. Chem.* **19**, 29 (1991).
177. C. T. Walsh and Y.-C. J. Chen, *Angew. Chem. Int. Ed. Engl.* **27**, 333 (1988).
178. O. Abril, C. C. Ryerson, C. Walsh, and G. M. Whitesides, *Bioorgan. Chem.* **17**, 41 (1989).
179. M. J. Taschner and D. J. Black, *J. Am. Chem. Soc.* **110**, 6892 (1988).
180. M. J. Taschner and Q.-Z. Chen, *Bioorg. Medicinal Chem. Lett.* **1**, 535 (1991).
181. V. Alphand, A. Archelas, and R. Furstoss, *J. Org. Chem.* **55**, 347 (1990).
182. G. Iacazio, G. Langrand, J. Baratti, G. Buono, and C. Triantaphylides, *J. Org. Chem.* **55**, 1690 (1990).
183. P. Zhang and K. S. Kyler, *J. Am. Chem. Soc.* **111**, 9241 (1989).
184. P. R. Bonneau, T. P. Graycar, D. A. Estell, and J. B. Jones, *J. Am. Chem. Soc.* **113**, 1026 (1991).
185. C. R. Dunn and co-workers, *Phil. Trans. R. Soc. Lond.* **332**, 177 (1991).
186. P. Wirsching, J. A. Ashley, S. J. Benkovic, K. D. Janda, and R. A. Lerner, *Science* **252**, 680 (1991).
187. R. A. Lerner, S. J. Benkovic, and P. G. Schultz, *Science* **252**, 659 (1991).

Aleksey Zaks
Schering-Plough Research Institute

EPINEPHRINE AND NOREPINEPHRINE

(-)-Epinephrine (E), (-)-norepinephrine (NE), and dopamine (DA) belong to a class of compounds known as catecholamines, and are neurotransmitters and hormones in peripheral tissues and the central nervous system. NE is the principal postganglionic sympathetic neurotransmitter in the periphery and in adrenergic neurons in the central nervous system. It is released from nerve terminals, and acts to transmit a nerve impulse across the synapse (synaptic cleft) between the nerve terminal and postsynaptic site of action. Many neurotransmitters, including E and NE, can be both inhibitory and excitatory depending on which tissue or organ the neurotransmitter is acting on. For example, NE may be excitatory, as in the heart and vasculature, inhibitory, as in the intestine, or both excitatory and inhibitory, as in the central nervous system depending on location.

The chromaffin cells of the adrenal medulla may be considered to be modified sympathetic neurons that are able to synthesize E from NE by *N*-methylation. In this case the amine is liberated into the circulation, where it exerts effects similar to those of NE; in addition, E exhibits effects different from those of NE, such as relaxation of lung muscle (hence its use in asthma). Small amounts of E are also found in the central nervous system, particularly in the brain stem where it may be involved in blood pressure regulation. DA, the precursor of NE, has biological activity in peripheral tissues such as the kidney, and serves as a neurotransmitter in several important pathways in the brain (1,2).

E occurs naturally in the adrenal medulla, and was one of the first chemical substances of biological importance to be isolated. In 1895 (3) extracts of adrenal glands were found to increase blood pressure and constrict smooth muscle. Numerous attempts were made to isolate the chemical compound that was responsible for the physiological effects, but the first positive findings were obtained in 1901 when E, together with small amounts of NE, was isolated and purified from bovine adrenal glands (4). Subsequently the structure of E was established (5). The chemical synthesis of the racemic form of E was published in 1904 (6). Although it was speculated after the discovery of E that another substance, probably NE, was released from stimulated sympathetic neurons (7,8), it was not until 1946 that it was demonstrated that mammalian peripheral sympathetic adrenergic neurons used NE as a transmitter instead of E (9). At that time it was not technically possible to study the content of transmitter in terminal parts of the adrenergic neuron. However, it was predicted that NE was concentrated in the nerve terminal region from which it was released to act as a neurotransmitter. This was conclusively demonstrated about 10 years later with the development of methods for visualizing catecholamines in freeze-dried tissues, and the demonstration that NE was released when sympathetic neurons such as the splenic nerve were stimulated (10).

Physical Properties

Some of the physical properties of E and NE are summarized in Table 1.

Table 1. Physical Properties of Epinephrine and Norepinephrine^a

Property	D-(+)-Epinephrine	D-(-)-Norepinephrine
IUPAC name	(-)-3,4-dihydroxy- α -[(methylamino)methyl]benzyl alcohol	(-)- α -(aminomethyl)-3,4-dihydroxybenzyl alcohol
CA name	(R)-(-)-4-[1-hydroxy-2-(methylamino)ethyl]-1,2-benzenediol	(R)-(-)-4-(2-amino-1-hydroxyethyl)-1,2-benzenediol
other names	epinephrine ^b , adrenaline	levarternol, D-arterenol, D-noradrenaline
CAS Registry Number	[51-43-4]	[51-41-2]
formula	C ₉ H ₁₃ NO ₃	C ₈ H ₁₁ NO ₃
formula wt	183.2	169.2
description appearance	colorless microcrystals ^c	colorless microcrystals
mp	ca 209–210°C (dec)	ca 212°C (dec)
solubility	in water, ca 0.1 g/100 mL; insoluble in alcohol and most other organic solvents	slightly soluble in water
absorption spectrum	0.1 M HCl λ_{max} 221 nm, ϵ ca 6100; λ_{max} 280 nm, ϵ ca 2700	in 0.1 M HCl, λ_{max} 221 nm, ϵ ca 5860; λ_{max} 279 nm, ϵ ca 2560
fluorescence spectrum	in 0.05 M sodium acetate buffer, pH 4, λ_{ex} 283 nm, λ_{em} 337 nm	in 0.05 M sodium acetate buffer, pH 4, λ_{ex} 283 nm, λ_{337} 337 nm; iodine oxidation in alkaline ascorbate ^d , λ_{ex} 412 nm, λ_{ex} 505 nm
reaction product		color reactions with various reagents ^e
fluorescence		
specific rotation	$[\alpha]_{\text{D}}^{25}$ -50 to -53.5°, <i>c</i> = 2 g/100 mL, 0.5 M HCl ^b ;	$[\alpha]_{\text{D}}^{25}$ 37.3°, <i>c</i> = 5 g/100 mL water containing 1 equiv hydrochloric acid ^f
optical purity	determined on the <i>N</i> -acetyl-3,4-di- <i>O</i> -acetyl derivative, $[\alpha]_{\text{D}}^{21}$ 94.7°, <i>c</i> = 1.01 g/100 mL, CHCl ₃ ^g	determined on the <i>N</i> -acetyl-3,4-di- <i>O</i> -acetyl derivative, $[\alpha]_{\text{D}}^{25}$ 81.3°, <i>c</i> = 1 g/100 mL, CHCl ₃ ^f resolution of (+)-norepinephrine ^h
source	resolution of (+)-epinephrine with (+)-tartaric acid	
likely impurities	(+)-epinephrine	(+)-norepinephrine

^a Data given in: *Specifications and Criteria for Biochemical Compounds, Supplement: Biogenic Amines and Related Compounds. Courtesy of the National Academy of Sciences.*

^b Ref. 11.

^c Acceptable preparations may be slightly off-white.

^d Ref. 12.

^e Ref. 13.

^f Ref. 14.

^g Ref. 15; report $[\alpha]_{\text{D}}^{20}$ -87.4°, *c* = 1 g/100 mL, CHCl₃ for the triacetyl derivative, and the configuration of natural (-)-epinephrine is the same as

D-()-mandelic acid.
h Ref. 16.

Stability. E is sensitive to air, light, heat, and alkalis. Metals, notably copper, iron, and zinc, destroy its activity. In solution with sulfite or bisulfite, it slowly forms an inactive sulfonate (17). The red color that forms when neutral or alkaline solutions are exposed to air is caused by adrenochrome.

NE is unstable in light and air, especially at neutral and alkaline pH. Oxidation to noradrenochrome occurs in the presence of oxygen and such divalent metal ions as copper, manganese, and nickel.

Analytical Methods

Bioassay methods were used in early studies as one of the more sensitive procedures for the estimation of E and NE in tissue extracts or biological fluids. However, the development of fluorescence methods (18) for the detection of nanomolar concentrations of E and NE in tissues, blood, and urine gradually displaced the original bioassays (19,20). This was followed by the development of radioenzymatic (21,22), gas chromatographic (gc) without (23) and with mass spectrometric detection (ms) (24). The latter, together with high performance liquid chromatographic (hplc) techniques, coupled with electrochemical detection (25), have replaced the earlier methods because of their increased sensitivity and specificity. In addition, their versatility and relative inexpensiveness as compared to gc–ms have gained widespread use and revolutionized the analytical procedures in this field. Improved techniques using capillary zone electrophoresis and combined hplc–ms have been developed, and will probably come into more general use (26).

Chemical Synthesis

The original commercial source of E was extraction from bovine adrenal glands (5). This was replaced by a synthetic route for E and NE (Fig. 1) similar to the original published route of synthesis (6). Friedel-Crafts acylation of catechol [120-80-9] with chloroacetyl chloride yields chloroacetocatechol [99-40-1]. Displacement of the chlorine by methylamine yields the methylamine derivative, adrenalone [99-45-6], which on catalytic reduction yields (±)-epinephrine [329-65-7]. Substitution of ammonia for methylamine in the sequence yields the amino derivative noradrenalone [499-61-6] which on reduction yields (±)-norepinephrine [138-65-8]. The racemic compounds were resolved with (+)-tartaric acid to give the physiologically active ()-enantiomers. The commercial synthesis of E and related compounds has been reviewed (27). The synthetic route for L-3,4-dihydroxyphenylalanine [59-92-7] (L-DOPA) has been described (28).

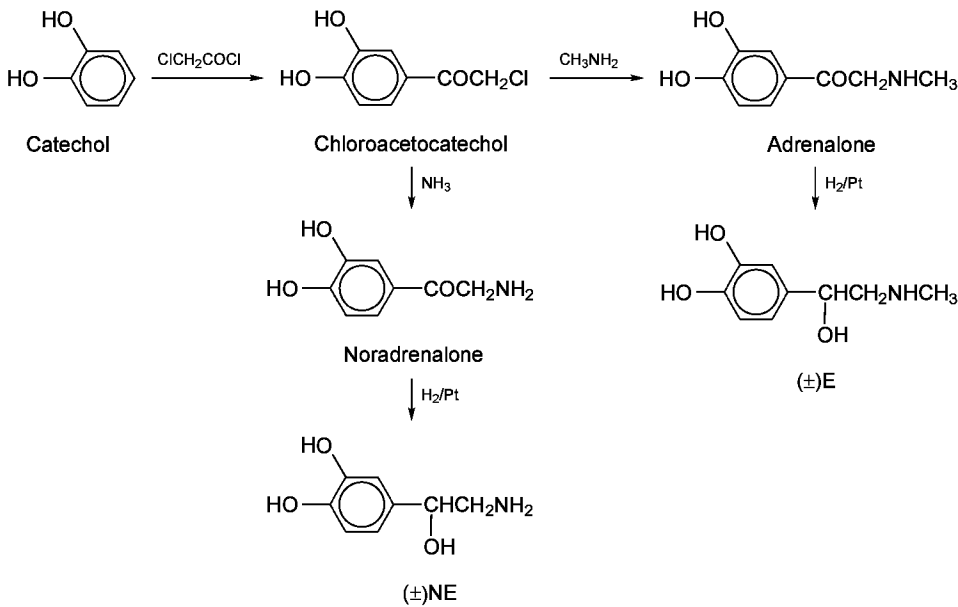


Fig. 1. Synthetic route to E and NE from catechol.

Economic Aspects

The principal trade names and suppliers of ()E, an adrenergic vasoconstrictor, are as follows: Adrenalin (Parke-Davis); Adrenalin in Oil (Parke-Davis); Primatene Mist (Whitehall); Bronkaid Mist (Sterling); and Epifrin (Allergan). The bitartrate salt [51-42-3] is used as an ophthalmic adrenergic vasoconstrictor; the principal trade names (suppliers) are Asmatane Mist (3M Riker); Epitrate (Wyeth-Ayerst); Medihaler-Epi (3M Riker); Primatene Mist Suspension (Whitehall); and Suprarenin (Sterling). ()Norepinephrine bitartrate [69815-49-2], used similarly to ()E, is supplied by Sterling under the trade name Levophed. Dopamine hydrochloride [62-31-7], a cardiovascular agent, is supplied under the name Dopestat from Parke-Davis and Intropin from Du Pont Pharmaceuticals. The total drugstore and hospital U.S. sales for 1991 were \$78,008,000 for ()E products, \$6,851,000 for ()NE products, and for dopamine \$14,595,000 (IMS Pharmaceutical Database Division). Levodopa [59-92-7], an antiparkinsonian agent, is supplied under the following trade names (manufacturers) as Bendopa (ICN), Dopan (Norwich-Eaton), Larodopa (Hoffmann-LaRoche), and Levopa (SmithKline-Beecham), and is a component of Madopa (Hoffmann-LaRoche) and Sinemet (Merck). The total U.S. drugstore and hospital sales for 1991 were \$63,535,000 with Sinemet controlling about 99% of this market (IMS Pharmaceutical Database Division).

Metabolism of Catecholamines

The general scheme of the biosynthesis of catecholamines was first postulated in 1939 (29) and finally confirmed in 1964 (Fig. 2) (30). Although not shown in Figure 2, in some cases the amino acid phenylalanine [63-91-2] can serve as a precursor; it is converted in the liver to (-)-tyrosine [60-18-4] by the enzyme phenylalanine hydroxylase. Four enzymes are involved in E formation in the adrenal medulla and certain neurons in the brain: tyrosine hydroxylase, dopa decarboxylase (also referred to as L-aromatic amino acid decarboxylase), dopamine-β-hydroxylase, and phenylethanolamine N-methyltransferase. Neurons that form DA as their transmitter lack the last two of these enzymes, and sympathetic neurons and other neurons in the central nervous system that form NE as a transmitter do not contain phenylethanolamine N-methyl-transferase. The component enzymes and their properties involved in the formation of catecholamines have been purified to homogeneity and their properties examined. The human genes for tyrosine hydroxylase, dopamine-β-oxidase and dopa decarboxylase, have been cloned (31,32). It is anticipated that further studies on the molecular structure and expression of these enzymes should yield interesting information about their regulation and function.

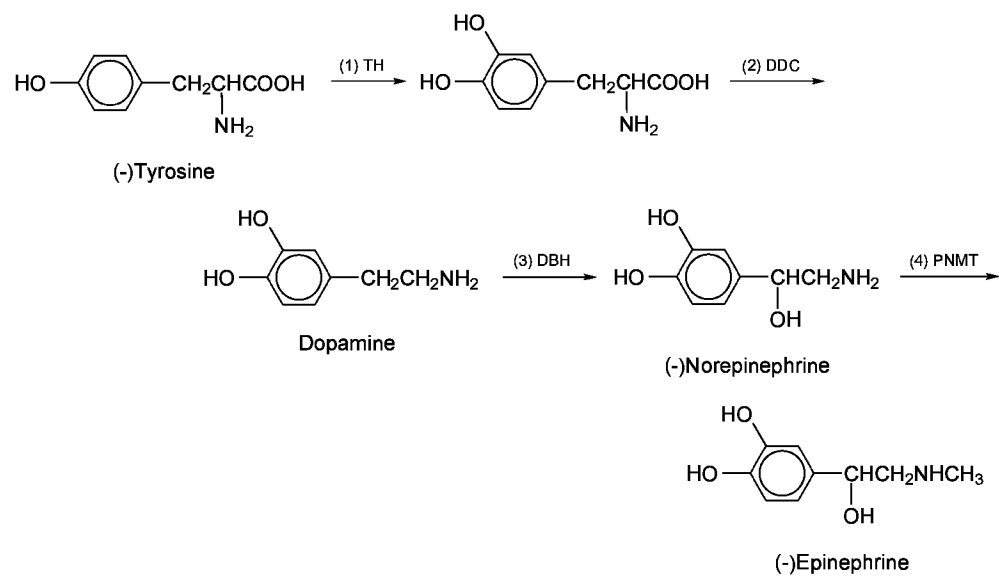


Fig. 2. Biosynthetic pathway for epinephrine, norepinephrine, and dopamine. The enzymes catalyzing the reaction are (1) tyrosine hydroxylase (TH), tetrahydrobiopterin and O₂ are also involved; (2) dopa decarboxylase (DDC) with pyridoxal phosphate; (3) dopamine-β-oxidase (DBH) with ascorbate, O₂ in the adrenal medulla, brain, and peripheral nerves; and (4) phenethanolamine *N*-methyltransferase (PNMT) with *S*-adenosylmethionine in the adrenal medulla and the brain.

An outline of the metabolism of E and NE is shown in Figure 3. By the action of monoamine oxidase (MAO) (33) and an aldehyde dehydrogenase, acid metabolites are formed, and by the action of catechol-*O*-methyltransferase (COMT) (34) 3-*O*-methylation of either the parent compounds or of the metabolites is affected. A listing of the catecholamine and metabolite abbreviations together with the CAS Registry Numbers is found in Table 2. The metabolism of DA leads to the production of 3,4-dihydroxyphenylacetic acid (DOPAC) and homovanillic acid (HVA) as shown in Figure 4. If 3-*O*-methylation first occurs, 3-methoxytyramine (3-MT) is formed which is then converted to 3-methoxy-4-hydroxyphenylethanol (MHPE). The aldehyde HAlD can also be oxidized to the carboxylic acid HVA. Catecholamines can also be conjugated to glucuronic and sulfuric acid (4).

Table 2. Catecholamines and Their Metabolites

Compound	Abbreviation	CAS Registry Number
dopamine	DA	[51-61-6]
(-)epinephrine (adrenaline)	E	[51-43-4]
(-)norepinephrine (noradrenaline)	NE	[51-41-2]
<i>Catecholamine metabolites</i>		
3,4-dihydroxymandelic acid (DHMA)	DOMA	[775-01-9] ^a
3,4-dihydroxymandelic aldehyde	D	[13023-73-9] ^a
3,4-dihydroxyphenylacetaldehyde	DAlD	[5707-55-1]
3,4-dihydroxyphenylacetic acid	DOPAC	[102-32-9]
3,4-dihydroxyphenylethanol	DOPET	[10597-60-1]
3,4-dihydroxyphenylethylene glycol	DOPEG	[3343-19-9] ^a
homovanillic acid (4-hydroxy-3-methoxy-phenylacetic acid)	HVA	[306-08-1]
4-hydroxy-3-methoxy-phenylacetaldehyde	HAlD	[5703-24-2]
metanephrine	M	[5001-33-2] ^a
3-methoxy-4-hydroxymandelic acid	3M4HMAld	[17592-23-3]
3-methoxy-4-hydroxyphenylethylene glycol	MHPG	[534-82-7] ^a
3-methoxy-4-hydroxyphenyl ethanol (MOPET)	MHPE	[2380-78-1]
3-methoxytyramine	3MT	[554-52-9]
normetanephrine	NM	[97-31-4] ^a
vanillylmandelic acid (4-hydroxy-3-methoxymandelic acid)	VMA	[55-10-7] ^a

^a Unspecified stereochemistry.

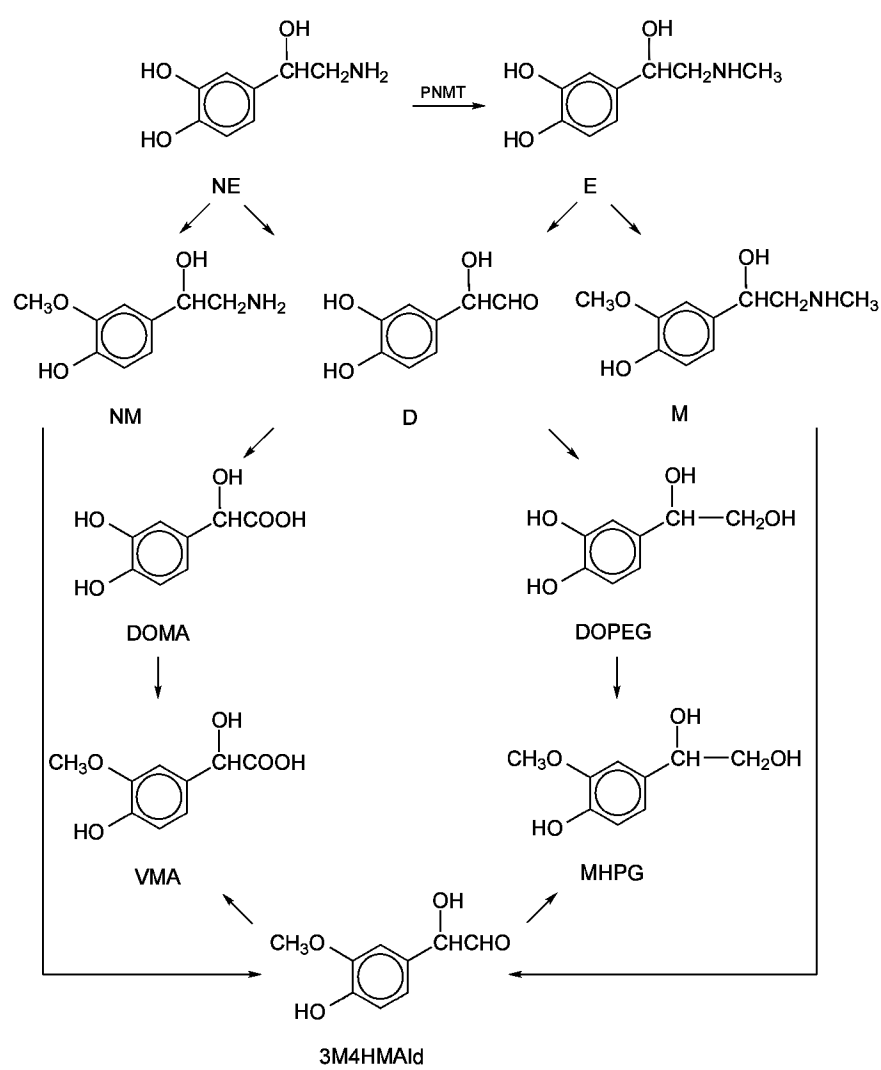


Fig. 3. Metabolism of norepinephrine and epinephrine (see Table 2 for abbreviations of amines and metabolites and CAS Registry Numbers).

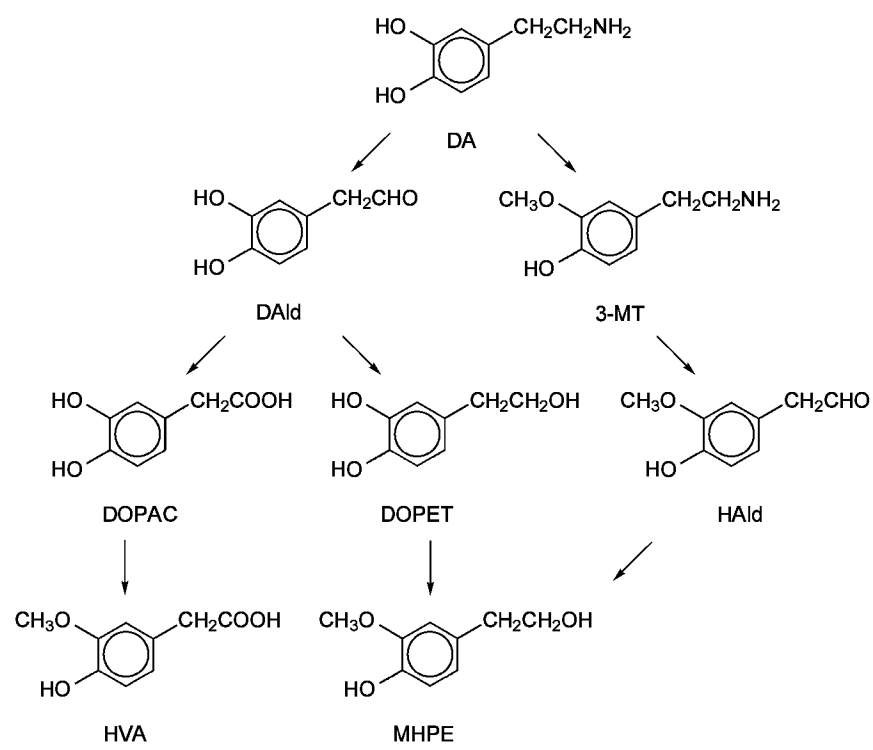


Fig. 4. Metabolism of dopamine (see Table 2 for abbreviations of amines and metabolites and CAS Registry Numbers).

Regulation of Synthesis, Storage, Release, and Metabolism

Catecholamine biosynthesis begins with the uptake of the amino acid tyrosine into the sympathetic neuronal cytoplasm, and conversion to DOPA by tyrosine hydroxylase. This enzyme is highly localized to the adrenal medulla, sympathetic nerves, and central adrenergic and dopaminergic nerves. Tyrosine hydroxylase activity is subject to feedback inhibition by its products DOPA, NE, and DA, and is the rate-limiting step in catecholamine synthesis; the enzyme can be blocked by the competitive inhibitor α -methyl-*p*-tyrosine (31). DOPA in the bloodstream can be taken up into neural tissue and into tissue devoid of tyrosine hydroxylase, thus bypassing the rate-limiting enzymatic synthetic step (35). Uptake of DOPA by the brain is the basis of the therapeutic effect of DOPA in the treatment of Parkinson's disease (a

disease characterized by depletion of DA in basal ganglia due to loss of nerve cells in the substantia nigra). In addition, the neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) (36) depletes DA in the striatum and produces a clinical syndrome like Parkinson's disease that responds to DOPA. DOPA in axonal cytoplasm is converted to DA by dopa decarboxylase. This enzyme is not specific for DOPA as its substrate, and is abundant in other neurons and tissues besides those containing catecholamines. The conversion of DOPA to DA outside the brain can be inhibited by carbidopa [28860-95-9], with the result that combined treatment with DOPA and carbidopa leads to greater brain DOPA levels than DOPA treatment alone.

DA in the cytoplasm of neurons is taken up into storage vesicles by an energy requiring process that can be inhibited by the alkaloid reserpine [50-55-5]. The depletion by reserpine is long lasting, and the storage granules are irreversibly damaged. Cytoplasmic NE can also be taken up stereoselectively into the vesicles. The vesicles contain dopamine-β-oxidase that catalyzes the conversion of DA to NE. In the vesicle NE can be stored, can leak back into the cytoplasm, or can be released into the synapse. Finally, at some sites, notably in adrenal medulla and brain stem, NE in *N*-methylated to form E by phenylethanolamine *N*-methyltransferase.

The storage vesicles play a dual role: they maintain a ready supply of catecholamines at the terminal (also in the adrenal medulla) available for release, and they mediate the process of release. For example, NE is liberated from nerve terminals of postganglionic sympathetic neurons (similarly E is released from adrenal medulla) when an electrical impulse or action potential propagates along the nerve and reaches the terminal and or adrenal medulla. The mechanism(s) regulating the release in the adrenal medulla or neurons are complex and not entirely known. However, the final step is thought to be similar. When an action potential reaches nerve terminals, calcium channels open, allowing an influx of the cation; increased intracellular calcium promotes fusion of the vesicles with the neuronal membrane and promotes the release into the extracellular space, by a process of exocytosis, of the soluble contents of vesicles, including dopamine-β-oxidase, NE, adenosine triphosphate, and in most cases peptides, as co-transmitters (37). Indirectly acting agents such as tyramine [51-67-2] and amphetamine [300-62-9] release catecholamines by a mechanism that is independent of calcium and is not associated with release of dopamine-β-oxidase, ie, nonexocytotically. These agents displace NE from storage vesicles, resulting in leakage from nerve terminals probably by a reversal of the usual inwardly directed process of neuronal uptake. These concepts have been reviewed (38).

Efficient regulatory mechanisms operate to modulate the rate of synthesis and release of catecholamines, depending on the need. For example, changes in blood pressure activate pressure-sensitive baroreceptors, sending signals back to the vasomotor centers in the brain. The latter then sends signals to the vasoconstrictor or vasodilator nerves, producing the appropriate alterations in activity by way of a decrease or increase in firing rate of neurons. At the nerve terminal level, such mechanisms as end-product inhibition of tyrosine hydroxylase, transient activation by phosphorylation of tyrosine hydroxylase, and enzyme induction all enable the neuron to respond to alterations in utilization of catecholamine neurotransmitters. Besides the latter, another important regulatory mechanism is the ability of released NE to interact with responsive sites, or receptors present not only on the postsynaptic neuron but also located presynaptically (in the case of NE, α-2 adrenergic receptors) on the neurons that are releasing the catecholamine to diminish NE release for a given amount of sympathetic nerve traffic. Thus, these presynaptic autoreceptors are believed to have the function of sensing the concentration of transmitter in the synaptic cleft and modulating the release and further synthesis of neurotransmitter accordingly. Other receptors such as muscarinic, dopamine, and opiate are also located presynaptically and are involved, for example, in adrenergic release. Similar presynaptic autoreceptor regulatory mechanisms appear to operate in the central nervous system for adrenergic and dopaminergic neurons (39).

Unlike the cholinergic system where the action of acetylcholine is terminated by cholinesterase, the action of catecholamine neurotransmitters released into the synaptic cleft is mainly brought to an end not by enzymatic degradation but by reuptake back into the nerve terminals (uptake 1) that released the catecholamines (40). Each of the three types of catecholamine neurons has its own uptake mechanism which is an energy-requiring, stereoselective process that can be blocked by certain classes of drugs such as tricyclic antidepressants (desipramine [50-47-5]), cocaine [50-36-2], and inhibitors of the Na–KATPase such as ouabain [630-60-4]. Nonneuronal uptake of catecholamines (uptake 2) also occurs, and is an energy-requiring, nonstereoselective process that can be blocked by adrenocorticosteroids such as corticosterone. A complementary DNA clone encoding a human NE transporter and more recently a DA transporter have been isolated.

Two important pathways for catecholamine metabolism are *O*-methylation by COMT, which is cytoplasmically localized, and oxidative deamination by the mitochondrial localized enzyme MAO. There are large amounts of MAO in tissues such as the liver and the heart which are responsible for the removal of most of the circulating monoamine, including some taken in from the diet. Tyramine is found in high concentrations in certain foods such as cheese, and in wine. Normally, this tyramine is deaminated in the liver. However, if MAO is inhibited, the tyramine may then be converted into octopamine [104-14-37] which may indirectly cause release of NE from nerve terminals to cause hypertensive crisis. Thus MAO, which is relatively nonspecific, plays an important role in the detoxification of pharmacologically active amines ingested from the diet.

MAO is known to occur in at least two forms, MAO A and MAO B, based on substrate selectivity, inhibition by various drugs, and cloning experiments. Clorgyline [17780-72-2] is a specific inhibitor of MAO A, which displays a substrate specificity for NE and serotonin. Deprenyl [2323-36-6] is a selective inhibitor of MAO B, and displays a substrate preference for β-phenylethylamine and benzylamine. Dopamine and tyramine are substrates for both enzymes.

Catecholamine Receptors

When catecholamines are released from adrenergic neuronal terminals or adrenal medulla, they are recognized by and interact with specific receptor proteins (adrenergic) on plasma membranes of effector cells or tissues that mediate the action of the agonist, resulting in a physiological response. The diverse actions of E and NE led to the 1948 proposal of the concept of distinct subtypes of adrenergic receptors, classified as alpha and beta adrenergic receptors, on the basis of the differential responses elicited by six sympathetic amines in various effector systems (41). Beta receptors were subsequently subdivided into beta-1 and beta-2 (42). In the 1970s, studies from a number of laboratories using functional (43) and radioligand-binding techniques (44) signaled the beginning of an exciting era of pharmacology in which subtypes of receptors could be more readily assessed. These studies revealed that alpha receptors could be further divided into alpha-1 and alpha-2 subpopulations (43). Similarly, it was proposed that the effects of DA could be accounted for by the existence of two DA receptor subtypes, DA D-1 and D-2. More recent developments from several experimental approaches, in particular molecular biology, have led to the cloning and expression of six structurally related alpha adrenergic receptors, three beta adrenergic receptors, and five, and very likely more, distinct types of dopamine receptor subtypes.

Pharmacological Actions

In general, activation of alpha-1 adrenergic receptors causes a contraction of smooth muscle and of blood vessels, pilomotor muscles, dilator pupillae, vas deferens, nictitating membrane, splenic capsule, and sphincters of the intestine and urinary bladder and of the bile duct. An exception is the relaxation of the smooth muscle of the intestine. Prazosin [19216-56-9], indoramin [26844-12-2], and WB-4101 are relatively selective antagonists of these receptors.

Alpha-2 adrenergic receptors are found in a wide variety of tissues from the central and peripheral nervous system, and also in adipocytes, blood platelets, pancreatic islets, and cultured cell lines. As expected for receptors with a varied distribution, alpha-2 adrenergic receptors are associated with many physiological functions, including regulation of blood vessels, locomotor activity, platelet aggregation, memory, anxiety, and sexual activity. Receptors of the alpha-2 variety are located presynaptically on the terminals of sympathetic neurons, where they mediate inhibition of neurotransmitter release, but they are also found in the central nervous system and postsynaptically at the periphery. Clonidine [4205-90-7] and other imidazoline analogues are selective agonists, and yohimbine [146-48-5] and idazoxan are relatively selective antagonists.

Beta receptors of the beta-1 subtype mediate an increase in heart rate and increased force of contraction; they are also found in the central nervous system. E and NE are equally potent agonists and selective antagonists are atenolol [29122-68-7] and betaxolol [63659-18-7]. Beta-2 receptors are well known for their involvement in relaxing bronchioles. E is a more potent agonist than NE; procaterol [72332-33-3] is a selective agonist; ICI 118551 and α-methylpropranolol are selective antagonists. A particular amine may act on both alpha and beta receptors or predominantly on one type. NE acts mainly on alpha-1, E on both alpha and beta, and isoproterenol [7683-59-2] almost exclusively on beta receptors. Numerous antagonists also differentiate between

alpha and beta receptors. The prototypic beta-adrenergic receptor antagonist propranolol [525-66-6] is essentially inactive at alpha receptors; the nonselective alpha-adrenergic receptor antagonist phentolamine is very weak or inactive at beta receptors.

Second Messengers

Adrenergic receptors comprise integral membrane proteins responsible for binding extracellular E and NE and initiating membrane signals through interaction with one of a series of guanine nucleotide binding regulatory proteins (*G*-proteins) to activate second messenger systems. In this respect there are two significant mechanisms involved in the interaction of E and NE with their receptors, the first involving changes in calcium conductance and the phosphoinositol (PI) response, and the second related to inhibition or activation of adenylyl cyclase.

Excitation of smooth muscle via alpha-1 receptors (eg, in the uterus, vascular smooth muscle) is accompanied by an increase in intracellular-free calcium, possibly by stimulation of phospholipase C which accelerates the breakdown of polyphosphoinositides to form the second messengers inositol triphosphate (IP₃) and diacylglycerol (DAG). IP₃ releases intracellular calcium, and DAG, by activation of protein kinase C, may also contribute to signal transduction. In addition, it is also thought that alpha-1 adrenergic receptors may be coupled to another second messenger, a pertussis toxin-sensitive *G*-protein that mediates the translocation of extracellular calcium.

Effects mediated through beta- and alpha-2 adrenergic receptors are related to activation or inhibition of adenylyl cyclase through coupling of the receptors to the inhibitory (*G_i*) or stimulatory (*G_s*) regulatory proteins that influence adenylyl cyclase through guanosine triphosphate. Beta receptor activation leads to stimulation of adenylyl cyclase, resulting in generation of the second messenger *c*-AMP, whereas alpha-2 activation leads to inhibition. In a similar manner, the effects of DA were originally thought to be mediated by only two distinct *G*-protein-coupled receptor subtypes, DA D-1, inducing stimulation, and DA-2, inducing either inhibition or no effect on adenylyl cyclase, respectively.

Molecular Biology

Extensive pharmacological data together with the emergence of molecular biological approaches in the study of receptor structure and function, have yielded new information concerning the classification of alpha- and beta-adrenergic and dopaminergic receptors. This new information has resulted in the identification of a number of distinct subtypes. The following subdivision of receptor subtypes is based on data generated on endogenous receptors and cloned receptors. Three distinct alpha-1 (designated 1A, 1B, and 1C), three distinct alpha-2 (designated 2A, 2B, and 2C), three beta (designated beta-1, beta-2, and beta-3), and five or six DA (designated D-1, ie, D-1A, D-1B, and D-1C, also known as D-5, and D-2, ie, D-2A [2L], D-2B [2L], D-3, and D-4) receptors have been characterized. The D-2 receptor exists as two isoforms (2A and 2B), which are generated through alternative splicing of *m*RNA and differ by the presence or absence of a 29-amino acid insert in the third cytoplasmic loop. The reader is referred to reviews concerning structure, characterization, and pharmacology of alpha (45,46), beta (47,48), and DA (49,50) receptor subtypes. The application of molecular biology to the study of adrenergic and dopaminergic receptors has occurred only in the last few years; much still has to be learned about the structure and function of these receptors, and the same for subtypes that have yet to be discovered.

A feature of the alpha and beta adrenergic and dopaminergic receptor subtypes is that they are members of a large superfamily of genes encoding numerous receptors that couple to guanine nucleotide regulatory proteins during signal transduction. Members of this family show common structural features, such as seven putative hydrophobic membrane spanning domains that are connected by hydrophilic loops that extend alternately into the extracellular and intracellular spaces from the plasma membrane. This model has the *N*-terminus outside the cell, with the protein weaving in and out of the membrane seven times with the C terminus in the cell. The seven putative transmembrane regions (about 175 amino acids) are relatively conserved among different family members (eg, the three alpha-2 adrenergic receptors share about 75° amino acid identity in their putative transmembrane regions) whereas the extra- and intracellular loop regions are more varied even between closely related receptors. Other similarities of this receptor protein family are that the sites for *N*-linked glycosylation are found in the *N*-terminus extracellular portion of the molecule, and there are a number of highly conserved amino acids within the transmembrane regions. Moreover, among *G*-protein-coupled receptors, a high degree of variability has been observed in the third cytoplasmic loop and the carboxy tail. Generally, the protein coding regions of the genes in this family are without introns, but there are exceptions including the alpha-1 adrenergic and D-2, D-3, and D-4 genes.

The primary correlates or effector pathways of the various receptor subtypes are summarized as follows: most alpha-1 adrenergic subtype mediated responses involve the stimulation of phosphoinositide metabolism and the generation of the second messengers IP₃ and DAG; alpha-2 adrenergic subtypes can inhibit adenylyl cyclase, acting through *G_i* resulting in a decrease in the second messenger, *c*-AMP, although some alpha-2 adrenergic mediated responses cannot be explained by an action on adenylate cyclase, and other signal transduction mechanisms such as Na⁺ H⁺ exchange, or calcium ion channels, or as yet unknown mechanisms may be involved, particularly with regard to alpha-2A and 2B subtypes. DA D-1, including D-1C (also known as D-5), and beta adrenergic subtypes, are linked to adenylyl cyclase and induce stimulation, resulting in an increase of *c*-AMP levels. DA D-2A and 2B receptors are characteristic of receptors coupled to inhibition of adenylyl cyclase inducing a decrease in *c*-AMP; other signal transduction mechanisms such as activation of potassium and inhibition of calcium channel activity may also be involved. No functional correlate has been described for D-3 and D-4 receptors. In many cases there are few compounds that exhibit high selectivity for any particular subtype; moreover, some of the subtypes (particularly adrenergic) have only been found in a certain tissue or species and therefore may be endogenous only to that tissue or species. Thus further studies will be required to find out whether all subtypes are found in humans and what is their localization and function.

Therapeutic Uses

As described, E, NE, DA, and related compounds exhibit a wide diversity of physiological functions. Patients in clinical shock are usually treated with an iv administered pressor amine, including isoproterenol (beta agonist selective) NE, E, DA, and dobutamine [34368-04-2]. The choice of agents depends on the circumstances responsible for the decrease in blood pressure and the low rate of tissue perfusion that occurs in shock. E is also the primary treatment for anaphylactic shock. E is commonly included in local anesthetic solutions to promote hemostasis, and by vasoconstriction to reduce absorption resulting in prolongation of anesthesia. Several related sympathomimetic vasoconstrictor amines (eg, phenylephrine hydrochloride [61-76-7]) are used for nasal congestion. Because of their relaxation of bronchial smooth muscle, E and selected beta-2 agonists are used to antagonize the bronchospasm observed in asthma. NE is used for treating hypotension during anesthesia when tissue perfusion is good.

Alpha-adrenergic blocking agents such as prazosin (alpha-1 selective), which causes vasodilation in both arteries and veins without usually causing reflex tachycardia, are used to treat mild to moderate hypertension. Nonselective beta-adrenergic antagonists such as propranolol are used in the treatment of hypertension (usually with a diuretic), as prophylaxis in angina pectoris, and for prophylaxis of supraventricular and ventricular arrhythmias and other selected disorders. Selective beta-1 adrenergic antagonists such as metoprolol [37350-58-6] are used mainly for the treatment of hypertension. In addition, clonidine (an alpha-2 agonist) and methyl dopa (metabolized to alpha-methylnorepinephrine in brain) act centrally on vasomotor centers of the brain to reduce sympathetic outflow to the peripheral vessels and thus are used, but to a lesser extent, in the treatment of hypertension.

In Parkinson's disease, treatment with the amine precursor DOPA (with the decarboxylase inhibitor carbidopa) (Fig. 2), has been shown to ameliorate the symptoms and signs of the condition and prolong life. Direct-acting dopamine agonists such as bromocriptine [25614-03-3] and pergolide [66104-22-1], plus the MAO B inhibitor deprenyl, which inhibits the breakdown of dopamine, are also being used, generally in conjunction with DOPA therapy in the treatment of Parkinson's. There are several other disorders of the central nervous system in which catecholamines have been shown to be involved and drugs that affect the actions of catecholamines have a therapeutic action. Dopamine receptor antagonists that encompass several chemical classes such as phenothiazines (eg, chlorpromazine [50-53-5], butyrophenones (eg, haloperidol [52-86-8]), and thioxanthene derivatives (eg, chlorprothixene [113-59-7]), are prescribed for the management of both acute and chronic psychoses and in nonpsychotic individuals who are delusional or excited (eg, mania). The therapeutic as well as some of the side effects (eg, extrapyramidal) are thought to be the result of blockade of dopaminergic neurotransmission in the limbic, nigrostriatal, and hypothalamic, mesocortical systems (51). In the treatment of depression, most antidepressants are believed to improve mood

by increasing catecholamine and or serotonin concentrations. Tricyclic antidepressants such as imipramine [50-49-7] and amitriptyline [50-48-6] potentiate the actions of amines presumably by blocking the inactivating re-uptake of these amines after release from the presynaptic neuron. More selective agents, such as the selective serotonin re-uptake blocker fluoxetine [54910-89-3] have become available. MAO inhibitors block the metabolism of catecholamines, and the subsequent rise in amine levels in the brain are thought to underlie the antidepressant effect. The effects of drugs in the treatment of depression have been reviewed (52).

Besides behavior and blood pressure, catecholamine neurons also have important roles in other brain functions. Regulation of neuroendocrine function is a well-known action of catecholamines; for example, DA agonists reduce serum prolactin concentration, especially in conditions of hypersecretion. Ingestive behavior can be modulated by brain catecholamines, and some appetite-suppressing drugs are believed to act via catecholaminergic influences. Catecholamines also participate in regulation of body temperature.

Toxicity

Untoward effects of both E and NE (usually to a lesser degree) are anxiety, headache, cerebral hemorrhage (from vasopressor effects), cardiac arrhythmias, especially in presence of digitalis and certain anesthetic agents, and pulmonary edema as a result of pulmonary hypertension. The minimum subcutaneous lethal dose of E is about 4 mg, but recoveries have occurred after accidental overdosage with 16 mg subcutaneously and 30 mg intravenously, followed by immediate supportive treatment.

BIBLIOGRAPHY

"Epinephrine" in *ECT* 1st ed., Vol. 5, pp. 763–768 by O. Kamm, Parke, Davis & Co.; in *ECT* 2nd ed., Vol. 8, pp. 231–238, by H. M. Crooks, Jr., Parke-Davis & Co.; "Epinephrine and Norepinephrine" in *ECT* 3rd ed., Vol. 9, pp. 241–250, by J. R. McLean, Warner-Lambert Parke-Davis.

1. P. B. Molinoff and J. Axelrod, *Ann. Rev. Biochem.* **40**, 465 (1971).
2. U. Trendelenburg and N. Weiner, eds., *Catecholamines, Handbook of Experimental Pharmacology*, Vol. 90 II, Springer-Verlag, Berlin, 1988, p. 427.
3. G. Oliver and E. Schaefer, *J. Physiol.* **18**, 230 (1895).
4. J. Takamine, *J. Soc. Chem. Ind.* **20**, 746 (1901); T. B. Aldrich, *Am. J. Physiol.* **5**, 457 (1901).
5. H. Pauly, *Chem. Berl.* **36**, 2944 (1903); H. A. D. Jowett, *J. Chem. Soc. Trans.* **85**, 192 (1904).
6. F. Stolz, *Chem. Berl.* **37**, 4149 (1904).
7. W. B. Cannon and J. E. Uridil, *Am. J. Physiol.* **58**, 353 (1921).
8. Z. M. Bacq, *Ann. Physiochem. Biol.* **10**, 467 (1934).
9. U. S. von Euler, *J. Physiol.* **105**, 38 (1946).
10. U. S. von Euler in H. Blaschko and E. Muscholl, eds., *Catecholamines, Handbuch der Experimentellen Pharmakologie*, Vol. 33, Springer-Verlag, Berlin, 1972, p. 186.
11. *The U.S. Pharmacopeia*, 22nd rev., Mack Printing Co., Easton, Pa., 1989, p. 500.
12. P. A. Shore and J. S. Olin, *J. Pharmacol. Exp. Ther.* **122**, 295 (1958).
13. C. F. Schwender in K. Florey, ed., *Analytical Profiles of Drug Substances*, Vol. 1, Academic Press, Inc., New York, 1972, p. 149.
14. L. H. Welsh, *J. Am. Pharm. Assoc. Sci. Ed.* **44**, 507 (1955).
15. L. H. Welsh, *J. Am. Chem. Soc.* **74**, 4967 (1952); P. Pratesi and co-workers, *J. Chem. Soc.* **Part 1–2**, 2069 (1958).
16. B. F. Tullar, *J. Am. Chem. Soc.* **76**, 2067 (1948); **U.S. Pat. 2,774,789** (Dec. 18, 1956), B. F. Tullar (to Stedding Drug Inc.).
17. L. C. Schroeter, T. Higuchi, and E. E. Schuler, *J. Am. Pharm. Sci.* **47**, 723 (1958).
18. A. H. Anton and D. F. Sayre, *J. Pharmacol. Exp. Ther.* **138**, 360 (1962).
19. S. Udenfriend, *Fluorescence Assay in Biology and Medicine*, Vol. 2, Academic Press, Inc., New York, 1969.
20. Y. Miura and M. Adachi in A. M. Krstulovic, ed., *Quantitative Analysis of Catecholamines and Related Compounds*, Ellis Horwood, Chichester, UK, 1986, p. 340.
21. P. G. Passon and J. D. Peuler, *Anal. Biochem.* **51**, 618 (1973).
22. M. G. Ziegler, L. C. Woodson, and B. Kennedy in Ref. 20, p. 263.
23. J.-M. Schmitter in Ref. 20, p. 24.
24. J. Roboz in Ref. 20, p. 46.
25. P. T. Kissinger and co-workers, *Clin. Chem.* **23**, 1449 (1977); B. Kagedal and D. S. Goldstein, *J. Chromatograph.* **429**, 177 (1988).
26. B. A. Davis, D. A. Durden, and A. A. Boulton, *J. Chromatograph.* **374**, 227 (1986).
27. K. R. Payne, *Ind. Chem.* **37**, 523 (1961); D. Lednicer and L. A. Metscher, *The Organic Chemistry of Drug Synthesis*, John Wiley & Sons, Inc., New York, 1977, p. 62.
28. **U.S. Pat. 3,714,242** (Jan. 30, 1973), G. M. Jaffe and W. R. Rehl (to Hoffmann-LaRoche Inc.).
29. H. Blaschko, *J. Physiol.* **96**, 50P (1939).
30. T. Nagatsu, M. Levitt, and S. Udenfriend, *J. Biol. Chem.* **239**, 2910 (1964).
31. J. R. Cooper, F. E. Bloom, and R. H. Roth, *The Biochemical Basis of Neuropharmacology*, Oxford University Press, New York, 1991, p. 224.
32. H. Ichinose and co-workers, *Biochem. Biophys. Res. Commun.* **164**, 1024 (1989).
33. K. F. Tipton, *Br. Med. Bull.* **29**, 116 (1973).
34. J. Axelrod, *Pharmacol. Rev.* **18**, 95 (1966).
35. D. S. Goldstein and co-workers, *J. Clin. Endocrinol. Metab.* **64**, 861 (1987).
36. A. Y. Deutch and co-workers, *Neurosci. Lett.* **68**, 51 (1986).
37. R. M. Weinshilboum and co-workers, *Science* **174**, 1349 (1971).
38. H. Winkler in U. Trendelenberg and N. Weiner, eds., *Handbook of Experimental Pharmacology*, Vol 90 I, Springer-Verlag, Berlin, 1988, p. 43.
39. S. Kalsner and T. C. Westfall, eds., *Presynaptic Receptors and the Question of Autoregulation of Neurotransmitter Release*, Vol. 604, Annals of the New York Academy of Science, New York, 1990.
40. J. Axelrod, *Science* **173**, 598 (1971).
41. R. P. Ahlquist, *Am. J. Physiol.* **153**, 586 (1948).
42. A. M. Lands and co-workers, *Nature* **214**, 597 (1967).
43. S. Berthelsen and W. A. Pettinger, *Life Sci.* **21**, 595 (1977).
44. H. I. Yamamura, S. J. Enna, and M. J. Kuhar, *Methods in Neurotransmitter Receptor Analysis*, Raven Press, Ltd., New York, 1990.
45. J. K. Harrison, W. R. Pearson, and K. R. Lynch, *Trends Pharmacol.* **12**, 62 (1991).
46. R. R. Ruffolo, Jr., and co-workers, *Pharmacol. Rev.* **43**, 475 (1991).
47. R. J. Lefkowitz, J. M. Stadel, and M. G. Caron, *Annu. Rev. Biochem.* **52**, 159 (1983).
48. L. Emorine and co-workers, *Science* **245**, 1118 (1989).
49. O. Civelli and co-workers, *Eur. J. Pharmacol., Mol. Pharmacol. Sect.* **207**, 277 (1991).
50. R. K. Sunahara and co-workers, *Nature* **350**, 614 (1991).
51. K. L. Davis and co-workers, *Am. J. Psychiatry* **148**, 1474 (1991).
52. W. Z. Potter, M. V. Rudorfer, and H. Manji, *N. Engl. J. Med.* **325**, 633 (1991).

Thomas A. Pugsley
Warner-Lambert/Parke-Davis

EPITAXY.

See ELECTRONIC MATERIALS; THIN FILMS, FILM FORMATION TECHNIQUES.

EPOXIDE POLYMERS.

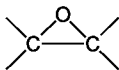
See EPOXY RESINS; POLYETHERS.

EPOXIDES.

See BUTYLENES; CHLOROHYDRINS; ETHYLENE OXIDE; OLEFINS; PROPYLENE OXIDE; STYRENES.

EPOXY RESINS

Epoxy resins are characterized by the presence of a three-membered ring known as the epoxy, epoxide, oxirane [75-21-8], or ethoxyline group. Commercial epoxy resins contain aliphatic, cycloaliphatic, or aromatic backbones. The capability of the epoxy ring to react with a variety of substrates imparts versatility to the resins. Treatment with curing agents gives insoluble and intractable thermoset polymers. In order to facilitate processing and modify cured resin properties, other constituents may be included in the compositions: fillers, solvents, diluents, plasticizers, accelerators, curatives, and tougheners.



Epoxy resins were first offered commercially in 1946 and are now used in a wide variety of industries. The principal use is in protective coatings, and the remainder in structural applications such as laminates and composites, tooling, molding, casting, construction, bonding and adhesives, and others. High chemical and corrosion resistance, good mechanical and thermal properties, outstanding adhesion to various substrates, low shrinkage upon cure, flexibility, good electrical properties, and ability to be processed under a variety of conditions are characteristics of epoxy resins.

The commercial possibilities for epoxy resins were first recognized by DeTrey Frères in Switzerland and DeVoe and Raynolds in the United States (1,2). In 1936, DeTrey Frères produced a low melting bisphenol A-based epoxy resin that gave a thermoset composition with phthalic anhydride. Application of the hardened composition was foreseen in dental products, but initial attempts to market the resin were unsuccessful. The patents were licensed to CIBA AG of Basel, Switzerland (now CIBA-GEIGY), and in 1946 the first epoxy adhesive was shown at the Swiss Industries Fair and samples of casting resin were offered to the electrical industry.

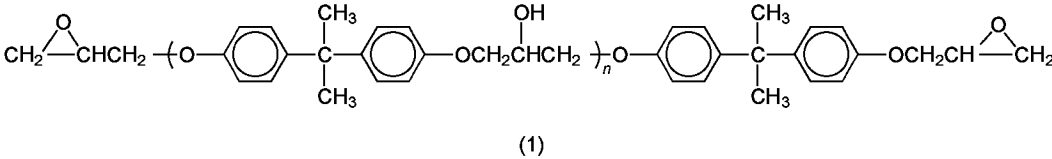
Immediately after World War II, DeVoe and Raynolds patented a series of epoxy resin compositions including esters. These resins were marketed through the subsidiary Jones-Dabney Co. Meanwhile, CIBA AG, under license from DeTrey Frères, further developed epoxy resins for casting, laminating, and adhesive applications, and the CIBA Products Co. was established in the United States. In the late 1940s, two U.S. companies, Shell Chemical Co. and Union Carbide Corp. (then Bakelite Co.), began research on bisphenol A-based epoxy resins. At that time, Shell was the only supplier of epichlorohydrin, and Bakelite was a leading supplier of phenolic resins and bisphenol A. In 1955, the four U.S. epoxy resin manufacturers entered into a cross-licensing agreement. Subsequently, The Dow Chemical Company and Reichhold Chemicals, Inc. joined the patent pool and began marketing epoxy resins.

In the 1960s, CIBA Products Co. marketed and manufactured glycidylated *o*-cresol novolak resins, which had been developed by Koppers Co. as high temperature-resistant polymers. Dow offered glycidylated phenol novolak resins, Shell introduced polyglycidyl ethers of tetrafunctional phenols, and Union Carbide developed a triglycidyl *p*-aminophenol resin.

The peracetic acid epoxidation of olefins was developed in the 1960s by Union Carbide in the United States and by CIBA AG in Europe for cycloaliphatic structures. CIBA Products marketed cycloaliphatic epoxy resins in 1963 and licensed several multifunctional resins from Union Carbide in 1965. The ensuing years witnessed the development of general-purpose epoxy resins with improved weathering characteristics based on the five-membered hydantoin ring and also on hydrogenated bisphenol A, but commercial success was limited and the products were withdrawn. Refinements in the manufacture of epoxy resins in the 1970s led to the development of electronic-grade materials with lower ionic impurities and development of resins with improved electrical properties. Increasing requirements in the composite industries in the 1980s led to the development of new multifunctional epoxy resins based on complex amine and phenolic structures with varying performance characteristics. More recently, there has been increased attention to the development of epoxy resins for water base applications using the latest formulating technologies.

Resin Properties

Epichlorohydrin and Bisphenol A-Derived Resins. The most widely used epoxy resins are diglycidyl ethers of bisphenol A [25068-38-6] (1) derived from bisphenol A [80-05-7] and epichlorohydrin [106-89-8].



The outstanding performance characteristics of the resins are conveyed by the bisphenol A moiety (toughness, rigidity, and elevated temperature performance), the ether linkages (chemical resistance), and the hydroxyl and epoxy groups (adhesive properties and formulation latitude, or reactivity with a wide variety of chemical curing agents) (see also PHENOLIC RESINS).

The bisphenol A-derived epoxy resins are most frequently cured with anhydrides, aliphatic amines, or polyamides, depending on desired end properties. Some of the outstanding properties are superior electrical properties, chemical resistance, heat resistance, and adhesion. Conventional epoxy resins range from low viscosity liquids to solid resins.

Diluents are commonly used to reduce the viscosity of epoxy systems to aid handling, improve ease of application, and to facilitate higher filler loading to reduce formulation cost. This, however, is achieved at the expense of other properties. To achieve a balance of properties, careful selection of diluent is needed. Table 1 qualitatively shows which diluent should be considered for minimal deterioration of properties.

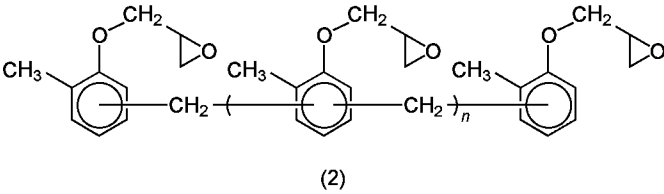
Table 1. Selectivity Chart of Reactive Diluents and Properties of System

Diluent	CAS Registry Number	Flexi-bil ity	Effi-cien cy	Solvent resistance	Acid resistance	Pot-life extension
butyl glycidyl ether	[2426-08-6]		x			
C ₈ -C ₁₀ -AME ^a	[68609-96-1]	x				
C ₁₂ -C ₁₄ -AME ^a	[68609-97-2]	x				x
cresyl glycidyl ether	[26447-14-3]				x	
neopentyl glycol diglycidyl ether	[17557-23-2]			x	x	

^a AME = aliphatic monoglycidyl ether.

Specialty Epoxy Resins. In addition to bisphenol, other polyols such as aliphatic glycols and novolaks are used to produce specialty resins. Epoxy resins may also include compounds based on aliphatic, cycloaliphatic, aromatic, and heterocyclic backbones. Glycidylation of active hydrogen-containing structures with epichlorohydrin and epoxidation of olefins with peracetic acid remain the important commercial procedures for introducing the oxirane group into various precursors of epoxy resins.

Epoxy Cresol–Novolak Resins (ECN). The cresol–novolak epoxy resins [37382-79-9] (2) are multifunctional, solid polymers characterized by low ionic and hydrolyzable chlorine impurities, high chemical resistance, and good thermal performance. ECN resins are widely used as base components in high performance electronic and structural molding compounds, high temperature adhesives, structural molding powders, castings and laminating systems, tooling applications, and powder coatings.



The epoxy cresol–novolak resins (2) are prepared by glycidylation of *o*-cresol–formaldehyde condensates in the same manner as the phenol–novolak resins. The *o*-cresol–formaldehyde condensates are prepared under acidic conditions with formaldehyde–*o*-cresol ratios of less than unity.

The multifunctionality contributes higher reactivity and cross-link density. These factors are especially critical when formulating systems that require improved thermal performance over conventional epichlorohydrin–bisphenol A systems. The melt viscosity of these resins, which are solids at room temperature, decreases sharply with increasing temperature. This affords the formulator an excellent tool for controlling flow of molding compounds, and facilitating the incorporation of ECN resins into other epoxies, eg, for powder coatings.

The *o*-cresol novolaks of commercial significance possess degrees of polymerization, *n*, of 1.7–4.4 and the epoxide functionality of the resultant glycidylated resins varies from 2.7 to 5.4. Softening points (Durrans') of the products are 35–99°C. The glycidylated phenol and *o*-cresol–novolak resins are soluble in ketones, 2-ethoxyethyl acetate, and toluene solvents. The commercial epoxy novolak products possess a residual hydrolyzable chlorine content of <0.15 wt% and a total chlorine content of ca 0.6 wt % (Table 2).

Table 2. Typical Properties of Epoxy Cresol–Novolak Resins^a

Property	ECN 1235	ECN 1273	ECN 1280	ECN 1299
mol wt	540	1080	1170	1270
epoxy value, eq 100 g	0.500	0.445	0.435	0.425
softening point, °C	34	73	80	99
density, kg L	1.11	1.16	1.17	1.19
epoxide functionality	2.7	4.8	5.1	5.4

^a Ref. 3.

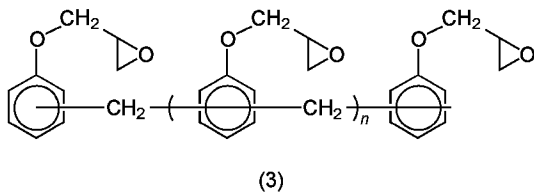
Bisphenol F Resin. Bisphenol F [2467-02-9] epoxy resin is of the same general structure as the epoxy phenol novolaks. Bisphenol F is 2,2'-methylene bisphenol. Whereas the epoxy phenol novolaks vary from viscous liquids to solid materials, the bisphenol F resin has a low viscosity (ca 4 Pa·s (40 P)) and 165 epoxy equivalent weight. Its *n* value (degree of polymerization) is about 0.15 and crystallization, often a problem with low viscosity conventional bisphenol A resins, is reduced with the bisphenol F resin.

Owing to relatively low viscosity, these resins offer advantages for 100% solids (solvent-free) systems. Higher filler levels are possible because of the low viscosity. Faster bubble release is also achieved. Higher epoxy content and functionality of bisphenol F epoxy resins can provide improved chemical resistance compared to conventional epoxies.

Bisphenol F epoxy resins are used in high-solids-high-build systems such as tank and pipe linings, industrial floors, road and bridge deck toppings, structural adhesives, grouts, coatings, and electrical varnishes. Bisphenol F epoxy resins are manufactured in Europe and Japan.

Epoxy Phenol–Novolak Resins. Epoxy phenol–novolak resins [9003-36-5] are represented by the general idealized structure (3) whereby multifunctional products are formed containing a phenolic hydroxyl group per phenyl ring in random para–para', ortho–para', and ortho–ortho' combinations.

Subsequent epoxidation with epichlorohydrin yields the highly functional epoxy novolak. The product can range from a high viscosity liquid of *n* = 0.2 to a solid of *n* value greater than 3.



The multi-epoxy functionality of the epoxy novolaks (2.2 to >5 epoxy groups per molecule) (3) produce more tightly cross-linked cured systems having improved elevated temperature performance and chemical resistance than the difunctional bisphenol A-based resins.

The thermal stability of epoxy phenol–novolac resins is useful in adhesives, structural and electrical laminates, coatings, castings, and encapsulations for elevated temperature service (Table 3). Filament-wound pipe and storage tanks, liners for pumps and other chemical process equipment, and corrosion-resistant coatings are typical applications using the chemically resistant properties of epoxy novolac resins.

Table 3. Properties of Epoxy Phenol Novolaks^a

Property	A ^a	B ^a	C ^a
<i>n</i> value	0.2	1.6	1.8
epoxide equivalent wt	175	178	200
epoxy functionality	2.2	3.6	3.8
viscosity ^b , mPa·s (–cP)	1,400	35,000	3,000 ^c
softening point, Durran, °C			53
color, Gardner	1	2	2
heat distortion temperature ^d , °C	156(165)	180(207)	182(202)

^a A D.E.N.431 (The Dow Chemical Company), EPN 1139 (CIBA-GEIGY Corp.); B D.E.N.438 (The Dow Chemical Company), EPN 1138 (CIBA-GEIGY Corp.); and C D.E.N.439 (The Dow Chemical Company).

^b At 52°C unless otherwise indicated.

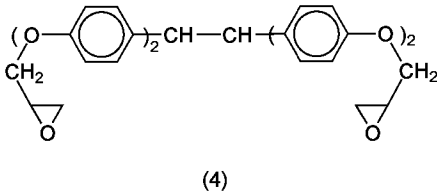
^c At 100°C.

^d Cured with methylenedianiline, gelled 16 h at 55°C + 2 h at 125°C + 2 h at 175°C (numbers in parentheses indicate additional 4 h at that temperature in °C).

Curing agents that give the optimum in elevated temperature properties for epoxy novolaks are those with good high temperature performance such as aromatic amines, catalytic curing agents, phenolics, and some anhydrides. When cured with polyamide or aliphatic polyamines and their adducts, epoxy novolac resins show improvement over bisphenol A-based epoxies, but the critical performance of each cure is limited by the performance of the curing agent.

Polynuclear Phenol–Glycidyl Ether-Derived Resins. This is one of the first commercially available polyfunctional products. Its polyfunctionality permits upgrading of thermal stability, chemical resistance, and electrical and mechanical properties of bisphenol A–epoxy systems. It is used in molding compounds and adhesives.

Tetrakis (4-hydroxyphenyl)ethane is prepared by reaction of glyoxal with phenol in the presence of HCl. The tetraglycidyl ether [27043-37-4] (4), mp ca 80°C, possesses a theoretical epoxide functionality of four with an epoxy equivalent weight of 185–208 (4).



Cycloaliphatic Epoxy Resins. This family of aliphatic, low viscosity epoxy resins consists of two principal varieties, cycloolefins epoxidized with peracetic acid and diglycidyl esters of cyclic dicarboxylic acids.

The cycloaliphatic products are generally liquids of lower viscosity than the standard glycidyl ether resins. The peroxidized resins contain no chlorine and low ash content and their ring-contained oxirane group (cyclohexene oxide type) reacts more readily with acidic curing agents than the bisphenol A-derived epoxy resins.

In addition, glycidyl esters are produced by the reaction of cycloaliphatic carboxylic acids with epichlorohydrin, followed by dehydrohalogenation with caustic. Such products are characterized by low viscosities (ca 500 mPa·s (–cP)). Reactivity of the glycidyl esters more closely resembles the standard bisphenol–epichlorohydrin resins.

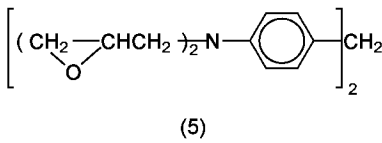
The nonaromatic nature of these materials provides for improved uv resistance and arc-track resistance compared to conventional epoxies. The best properties are generally achieved with anhydride and phenolic curing agents. These cycloaliphatic epoxy resins respond sluggishly to amine curing agents compared to standard bisphenol A-derived products. The reactivity of cycloaliphatic resins in curing reactions is discussed in Reference 5. The cyclohexane ring structure and short side chains give rise to high deflection temperatures but rather brittle products.

Recommended applications include transformers, insulators, bushings, wire and cable, generators, motors and switchgear, additives for adhesives, vinyl stabilization, and as viscosity depressants.

Aromatic and Heterocyclic Glycidyl Amine Resins. Among the specialty epoxy resins containing an aromatic amine backbone, the following are commercially significant.

Tetraglycidylmethylenedianiline-Derived Resins. Resins from aromatic glycidyl amines can be formulated into hot-melt or solution-binder systems with various reinforcements, eg, glass, graphite, boron, or aramid. These systems provide for simple cure cycles with excellent mechanical and adhesive properties at room and elevated temperatures. They are utilized for graphite-reinforced composites in aerospace and leisure products, structural adhesives, laminates, tooling and casting applications, and structures such as wings and fuselages (see ABLATIVE MATERIALS; ADHESIVES; COMPOSITE MATERIALS).

N,N,N',N'-Tetraglycidyl-4,4'-diaminodiphenylmethane [28768-32-3] (5) has an epoxy equivalent weight of 117–133, and a viscosity of 25 Pa·s (250 P) at 50°C (6).



Triglycidyl *p*-Aminophenol-Derived Resins. Resins derived from triglycidyl *p*-aminophenol [5026-74-7] originally developed by Union Carbide Corp. (7) are currently marketed by CIBA-GEIGY. Synthesis is conducted by reaction of epichlorohydrin with the phenolic and amino groups followed by dehydrohalogenation. The product is a viscous liquid (1.5–5 Pa·s (15–50 P) at 25°C) which is considerably more reactive toward amines than standard bisphenol A-derived resins. The epoxy equivalent weight is 105–114.

This low viscosity resin permits cure at low (70°C) temperatures and rapidly develops excellent elevated temperature properties. Used to increase heat resistance and cure speed of bisphenol A epoxy resins, it has utility in such diverse applications as adhesives, tooling compounds, and laminating systems. A molecularly distilled version is used as a binder for solid propellants (see EXPLOSIVES AND PROPELLANTS) and for military flares (see PYROTECHNICS). Its chief uses depend on properties of low viscosity and low temperature reactivity, particularly with carboxy-terminated rubbers.

Triazine-Based Resin. Triglycidyl isocyanurate [2451-62-9] (6) is a solid resin that provides superior thermal, electrical, and mechanical properties and is recommended for laminates, insulating varnishes, coatings, and adhesives. Widely used as a curing agent for special polyester-based weatherable powder coatings, it is also used in electronic applications owing to its retention of optical transparency after aging at temperatures up to 150°C and minimal smoke evolution on thermal decomposition (see EMBEDDING).

The triazine ring-containing product 1,3,5-triglycidyl isocyanurate (6) is synthesized by glycidylation of cyanuric acid with epichlorohydrin. The commercial product is a crystalline powder that exhibits an epoxy equivalent weight of ca 108 and softens in the 85–110°C range (see CYANURIC AND ISOCYANURIC ACIDS).

A summary of typical properties for multifunctional resins is shown in Table 4.

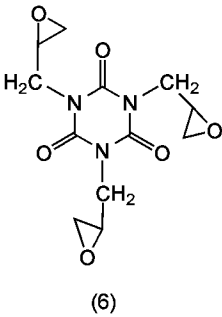


Table 4. Typical Properties of Multifunctional Epoxy Resins

Property	Amine-based		Other	
	MY 0500 ^a	MY 720 ^b	MT 0163 ^c	PT 810 ^d
physical form	liquid	semisolid	solid	solid
viscosity at 25°C, Pa·s ^e	1.5–5.0	10–35 ^f		
epoxy value, equiv 100 g, min	0.86	0.75	0.38	1.05
functionality	3	4	4	3

^a Triglycidyl derivative of *p*-aminophenol (7).

^b Tetraglycidyl derivative of methylenedianiline (5) (6).

^c Polynuclear phenol glycidyl ether (4) (4).

^d Triglycidyl isocyanurate (6) (8).

^e To convert Pa·s to Poise, multiply by 10.

^f Viscosity measured at 50°C.

Resin Synthesis and Manufacture

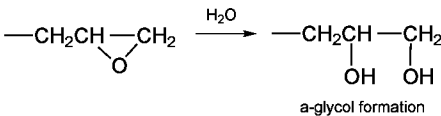
Epichlorohydrin and Bisphenol A-Derived Resins. Liquid epoxy resins may be synthesized by a two-step reaction of an excess of epichlorohydrin to bisphenol A in the presence of an alkaline catalyst. The reaction consists initially in the formation of the dichlorohydrin of bisphenol A and further reaction by dehydrohalogenation of the intermediate product with a stoichiometric quantity of alkali.

The pure diglycidyl ether of bisphenol A [1675-54-3] DGEBA is a crystalline solid (mp 43°C) with a weight per epoxide (WPE) = 170. The typical commercial unmodified liquid resins are viscous liquids with viscosities of 11–16 Pa·s (110–160 P) at 25°C, and an epoxy equivalent weight of ca 188.

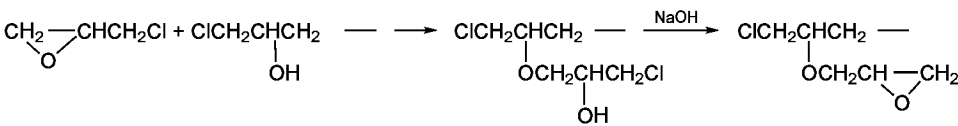
In the preparation of commercial DGEBPA, an excess of epichlorohydrin is used in order to minimize polymerization of the reactants to higher molecular-weight species. Nevertheless, the typical viscous final product usually contains ca 80% by weight of the monomeric (*n* = 0) DGEBPA as determined by gel-permeation chromatography (gpc). The manufacture of liquid epoxy resins in a batch process has been described in some detail (9).

The functionality of the terminal epoxy groups approaches values greater than 1.9, but this is in large measure a function of controlling the following side-reactions:

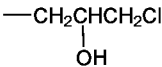
(1) Hydrolysis of epoxy groups:



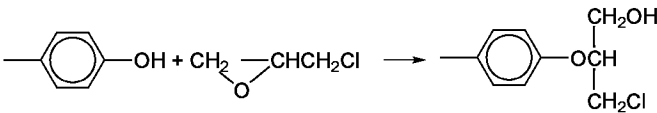
(2) Formation of bound chlorine by reaction of epichlorohydrin with hydroxy groups in the polymer backbones:



The bound chlorine formed does not readily saponify with metal hydroxide solutions and is analyzed as part of the total chlorine of the resin.
(3) Incomplete dehydrohalogenation which results in residual saponifiable or hydrolyzable chlorine:



(4) Abnormal addition of epichlorohydrin, ie, addition to the more substituted carbon atom:



In recent years, production of liquid resins of higher purity, ie, higher monomer content and fewer side-reactions, has been accomplished. This is in response to more stringent product quality requirements.
Aliphatic Glycidyl Ethers. Aliphatic epoxy resins have been synthesized by glycidylation of difunctional or polyfunctional polyols such as a 1,4-butanediol, 2,2-dimethyl-1,3-propanediol (neopentyl glycol), polypropylene glycols, glycerol, trimethylolpropane, and pentaerythritol.
The epoxidation is generally conducted in two steps: (1) the polyol is added to epichlorohydrin in the presence of a Lewis acid catalyst (stannic chloride, boron trifluoride) to produce the chlorohydrin intermediate, and (2) the intermediate is dehydrohalogenated with sodium hydroxide to yield the aliphatic glycidyl ether. A prominent side-reaction is the conversion of aliphatic hydroxyl groups (formed by the initial reaction) into chloromethyl groups by epichlorohydrin. The aliphatic glycidyl ether resins are used as flexibilizers for aromatic resins and as reactive diluents to reduce viscosities in resin systems.

Monofunctional aliphatic glycidyl ethers, eg, based on *n*-butanol or mixed C₈–C₁₄ alcohols, are used exclusively as reactive diluents to reduce viscosities of epoxy resin systems. Some loss of desirable cured properties results from the lowered functionality of the systems.
Polyfunctional aliphatic resins have exhibited high reactivity and degrees of cure with amines but problems of toxicity have diminished their usefulness and commercial interest. Solid epoxy resins can be prepared by the taffy process or the advancement process.

Taffy Process. Bisphenol A reacts directly with epichlorohydrin in the presence of a stoichiometric amount of caustic. The molecular weight of the product is governed by the ratio of epichlorohydrin–bisphenol A.

In practice, the taffy process is generally employed for only medium molecular-weight resins (1) (*n* = 1–4). The polymerization reaction results in a highly viscous product (emulsion of water and resin) and the condensation reaction becomes dependent on agitation. At the completion of the reaction, the heterogeneous mixture consists of an alkaline brine solution and a water–resin emulsion and recovery of the product is accomplished by separation of phases, washing of the taffy resin with water, and removal of water under vacuum.

Advancement Process. In the advancement process, sometimes referred to as the fusion method, liquid epoxy resin (crude diglycidyl ether of bisphenol A) is chain-extended with bisphenol A in the presence of a catalyst to yield higher polymerized products. The advancement reaction is conducted at elevated temperatures (175–200°C) and is monitored for epoxy value and viscosity specifications. The finished product is isolated by cooling and crushing or flaking the molten resin or by allowing it to solidify in containers.

The advancement process is more widely used in commercial practice. Isolation of the polymerized product is simpler because removal of copious amounts of NaCl is unnecessary.

The degree of polymerization is dictated by the ratio of liquid resin (crude DGEBA) to bisphenol A; an excess of the former provides epoxy terminal groups. The actual molecular weights attained depend on the purity of the starting material. Reactive monofunctional groups act as chain terminators.

Solid epoxy resins are sometimes designated as 1-, 4-, 7-, or 9-type resins; these approximate the degree of polymerization. Commercial products are designated similarly, eg, Epon 1001, 1004, 1007, and 1009 (Shell Chemical Co.). The relationship between *n* value, epoxy equivalent weight, and melting point is shown in Table 5.

Table 5. Properties of Commercial Epoxy Resins

Average <i>n</i> value	Epoxy equivalent wt	Mp ^a , °C	Approximate mol wt
0.1	185	liquid	360–380
0.3	210	liquid	390–450
0.6	255	liquid	460–560
2.2	487	65–75	850–1100
5.5	950	95–105	1750–2050
14.4	2250	125–135	4000–5000
16.0	3250	145–155	5000–8000

^a Duran's.

Calculation of the amount of bisphenol A for a typical advancement to a commercial grade of solid epoxy resin is based on the epoxy value of the starting material and the epoxy value desired:

bisphenol A charge = $\frac{1000(E_i - E_f)}{8.77 + E_f}$

where *E_i* refers to the epoxy value of the liquid resin expressed in equivalents per 1000 g using a charge of 1000 g of liquid resin; and *E_f* refers to the final epoxy value of the desired solid resin product. Thus, for a 7-type solid resin with a *E_f* of 0.55 eq / 1000 g as the target epoxy value, the bisphenol A charge is computed as follows:

bisphenol A charge = $\frac{1000(5.30-0.55)}{8.77 + 0.55} = \frac{4750}{9.32}$
= 509.6 g (per 1000 g liquid resin)

As the molecular weight of the resin increases (greater *n* value), a corresponding reduction in the terminal epoxy content of the molecule occurs with an increase in the hydroxyl content. This is accompanied by an increase in viscosity and melting point and eventually leads to resins with sufficiently high molecular weight such that the epoxy content is negligible. The resins are thus linear polyethers with pendent hydroxyl groups. Such resins are usually supplied in organic solvents and curing of the polyols with hydroxyl-reactive agents is usually required to develop their properties. A family of higher molecular-weight thermoplastic solid and solution resins (Union Carbide) is a related class of polymers or engineering plastics (Phenoxies).

Gel-permeation chromatography studies of epoxy resins prepared by the taffy process shown *n* values 0, 1, 2, 3, etc, whereas only even-numbered repeat units are observed for resins prepared by the advancement process. This is a consequence of adding a difunctional phenol to a diglycidyl ether derivative of a difunctional phenol in the polymer-forming step.

In recent years, proprietary catalysts for advancement have been incorporated in precatalyzed liquid resins. Thus only the addition of bisphenol A is needed to produce solid epoxy resins. Use of the catalysts is claimed to provide resins free from branching which can occur in conventional fusion processes (10). Additionally, use of the catalysts results in rapid chain-extension reactions because of the high amount of heat generated in the processing.

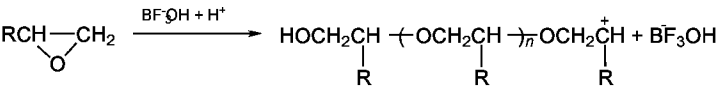
The preparation of flame-retardant epoxy resins is accompanied by inclusion of tetrabromobisphenol A [79-94-7] in the advancement process (see FLAME RETARDANTS). Products containing ca 20 wt % Br are extensively employed in the printed circuit board industry.

Liquid resins containing bromine (ca 49 wt %) can also be prepared directly from tetrabromobisphenol A and epichlorohydrin and are used for critical applications where a high degree of flame retardancy is required.

Curing Reactions

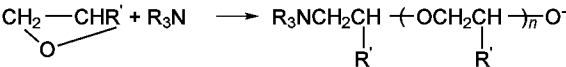
A variety of reagents has been described for converting the liquid and solid epoxy resins to the cured state, which is necessary for the development of the inherent properties of the resins. Liquid epoxy resins contain mainly epoxy groups and solid resins are composed of both epoxy and hydroxyl curing sites. The curing agents or hardeners are categorized as either catalytic or coreactive and the functional groups of the resins are terminal epoxy together with a pendent hydroxyl per repeat unit of the polymer chain.

Catalytic curing agents initiate resin homopolymerization, either cationic or anionic, as a consequence of using a Lewis acid or base in the curing process. The Lewis acid catalysts frequently employed are complexes of boron trifluoride with amines or ethers.



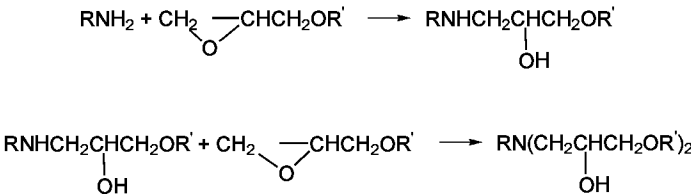
Mechanisms for polymerization of epoxides by Lewis acids are proposed in References 11–15.

The most important Lewis bases are tertiary amines or polyamines converted into tertiary amines upon reaction with epoxide groups.



Mechanisms for reaction of tertiary amines with epoxides are discussed in References 16 and 17.

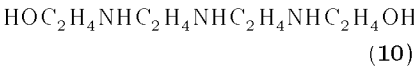
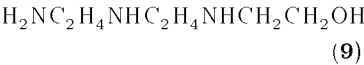
Coreactive curing agents are polyfunctional reagents that are employed in stoichiometric quantities with epoxy resins and possess active hydrogen atoms. The important classes include polyamines, polyamides (formed from polyamines and dimerized fatty acids), polyphenols, polymeric thiols, polycarboxylic acids, and anhydrides. Polyamines constitute a large class of hardeners with aliphatic, aromatic, cycloaliphatic, and heterocyclic groups. Cure of liquid epoxy resins is readily accomplished via the epoxy groups at room temperature with nonaromatic amines and at slightly elevated temperatures with aromatic amines, although reaction of the latter can be accelerated to function at RT. The mechanism of the reaction between epoxy resins and primary aliphatic amines is discussed in detail in References 18–23. The overall curing reaction may be depicted as follows:



Detailed discussions of network formation from amines and epoxy resins are provided in References 24 and 25.

Polyamides provide RT cure of epoxy-terminated resins as well as flexibilization; they are derived by reaction of dimerized vegetable oil fatty acids (dimer acids) with polyamines.

Comparisons of the properties of solvent-free liquid coatings based on unmodified bisphenol A epoxies cured with varying amine hardeners are shown in Table 6. Examples include diethylenetriamine [111-40-0], H₂NC₂H₄NHC₂H₄NH₂ (7), triethylenetetramine [112-24-3], H₂NC₂H₄NHC₂H₄NHC₂H₄NH₂ (8), *N*-(2-hydroxyethyl)diethylenetriamine [1965-29-3] (9), and *N,N'*-di(2-hydroxyethyl)diethylenetriamine [4484-60-0].



For solvent-based ambient cure systems polyamides are often the hardeners of choice. For heat-cured systems, anhydrides are used to provide higher heat-resistance systems but at the expense of flexibility.

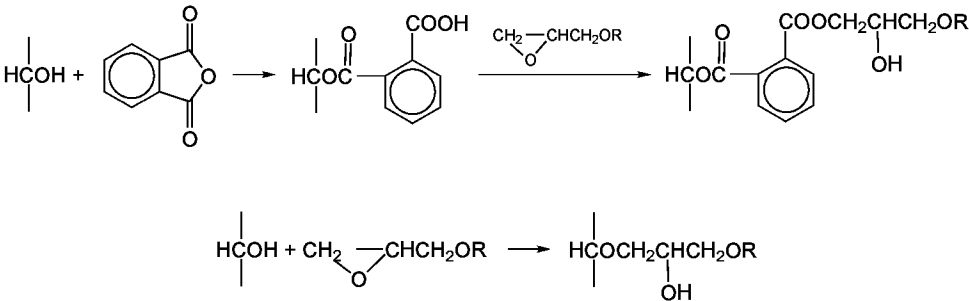
Table 6. Comparative Properties of Bisphenol A–Epoxy Resins Cured with Different Hardeners^a

Properties	Short-chain aliphatic polyamine	Oxyalkylated short-chain polyamine	Long-chain polyamine adduct	Aromatic polyamine adduct	Polyamide
chemical structure	(7), (8)	(9), (10)	trimethyl-1,6-hexan ediamine [25620-58-0]		reaction product of polyamines with dimer acids
advantages	low viscosity, generally good chemical resistance	low viscosity, generally good chemical resistance,	color stability, generally good chemical resistance,	long pot life, good acid resistance, cures in presence of	long pot life, generally good solvent resistance,

		low toxicity	good flexibility	moisture	good flexibility and adhesion
limitations	dermatitis potential, poor organic acid resistance, mixing ratio critical, film appearance	short pot life, poor organic acid resistance	short pot life, low organic acid resistance	high viscosity, color, low resistance to aromatic solvents	poor organic acid resistance, blushing
application-handling	100:10 or 13	100:25	100:35	100:60	100:43
mixing ratio by wt					
viscosity, mPa·s (cP)	4,500	3,000	7,800	8,000	>10,000
pot life, min	30–60	<30	23	135	75
color, Gardner	1–4	5–8	1–2	>12	6
dry time (paper-free), h	1.5–3	<10	ca 2	ca 16	
film quality					
flow	poor	poor	good	good	good
surface	blushing	blushing	blush-resistant	blush-resistant	blushing
color stability	good	good	good	poor	fair

^a Bais: resin diglycidyl ether of bisphenol A (wt per epoxide (WPE) 182–196).
Courtesy of the Federation of Societies for Coatings Technology.

Polyphenols or phenol-terminated resins are utilized to effect chemical cross-linking of epoxy resins with added catalysts or accelerators for the reaction (26).
The reactions of carboxylic acids and anhydrides with epoxy resins have been extensively studied in a variety of investigations, particularly References 27–31. The general reaction of epoxide resins and anhydrides is



Typical epoxy–anhydride systems are described in Table 7.

Table 7. Epoxy–Anhydride System Handling and Performance

Anhydride name	CAS Registry Number	Melting point or viscosity at 25°C	Mol wt	Recommended concentration, phr ^a	Curing cycles	ASTM deflection temperature, °C
phthalic anhydride	[84-44-9]	131°C	148	30–75	4 to 24 h at 150°C	110–147
methyl-4-endomethylene-tetrahydrophthalic anhydride	[129-64-6]	175–275 mPa·s ^b	178	80–90	2 h at 140°C + 2–192 h at 200°C	150–175
hexahydrophthalic anhydride	[85-42-7]	35–37°C	154	50–100	2 h at 100°C, or 2 h at 100°C + 2 h at 149°C	110–130
tetrahydrophthalic anhydride	[85-43-8]	99–101°C	152	70–80	2 h at 100°C + 4 h at 150°C	122
dodecenylsuccinic anhydride	[25377-73-5]	290 mPa·s ^b	266	130–150	2 h at 121°C or 1 h at 100°C + 2 h at 204°C	60–70

^a phr Parts per hundred parts of resin.
^b mPa·s cP.

Epoxy resins contain both epoxy terminal groups and pendent hydroxyl groups and are cured in accordance with available sites present in a given resin. Low and intermediate weight resins respond to curing by reagents such as dicyandiamide (see CYANAMIDES) and imidazoles at elevated temperatures. The curing reactions of these reagents are discussed in References 32–36.
Cross-linking with aminoplasts and phenoplasts constitutes an important class of hardeners for high molecular-weight epoxy resins that require elevated temperature cures (see AMINO RESINS).

Economic Aspects

Epoxy resin sales increased rapidly in the 1970s and continued into the 1980s as new applications were developed. Worldwide consumption of epoxy resins is shown in Table 8 where the proportions of the primary applications are shown.

Table 8. 1989 Consumption of Epoxy Resins by Region, ^a 10³ t

Resin	United States	Western Europe	Japan	Total
coatings	90	99	52	241
electrical electronic ^b	40	32	55	127

composites	12	13		25
construction ^c	11	31	13	55
adhesives	12	8	7	27
other ^d	19	4	10	33
<i>Total</i>	<i>184</i>	<i>187</i>	<i>137</i>	<i>508</i>

^a Data used with permission (37).
^b Includes tooling applications in the United States.
^c Significant amounts of coatings are included in this category for Western Europe.
^d Includes composites applications in Japan.

High performance multifunctional epoxy resins have achieved a higher compounded growth rate than the significantly larger volume DGEBPA conventional resins. Although epoxy resins cost more than competitive products, their longer service life and high performance capabilities provide a better cost-performance ratio.

The fastest growth over the past decade has been in laminates and composites. Tremendous growth in the electronics market has markedly increased the demand for epoxy laminating resins for the manufacture of printed wiring boards and epoxy-molding compounds for semiconductor encapsulation. Use of epoxy composites in the transportation industry, including the military and commercial aircraft and automotive fields, where a high strength-to-weight ratio is required, is growing at a steady rate that is expected to increase through the end of the century (see COMPOSITE MATERIALS, POLYMER MATRIX). U.S. production and sales of epoxy resins of all types are shown in Table 9 for the years 1980–1989.

Table 9. U.S. Production and Sales of Epoxy Resins^{a, b} 10³ t

Year	Production	Domestic sales and use	Total sales and use
1980	143	129	145
1981	153	132	151
1982	130	119	137
1983	152	134	150
1984	185	154	172
1985	175	148	167
1986	181	145	171
1987	197	156	190
1988	221	171	213
1989	232	180	219

^a Data used with permission (37).
^b Ref. 38.

Selling prices in the United States of various important epoxy resin products are listed in Table 10. As with other petrochemical-based products, prices of epoxies have risen rapidly during the 1980s and 1990s as oil prices have escalated.

Table 10. Average U.S. Unit Sales Values for Epoxy Resins,^{a, b} \$/kg

Year	Unmodified resins	Modified advanced resins
1980	2.31	2.80
1981	2.60	2.93
1982	2.73	2.98
1983	2.71	2.98
1984	2.89	3.17
1985	2.67	3.06
1986	2.40	3.04
1987	2.45	3.35
1988	2.25	2.98

^a Data used with permission (37).
^b Ref. 39.

Trademarks of the principal epoxy resin producers are as follows:

Company	Trade name
Shell Chemical Co.	EPON, EPONOL, Epikote
Dow Chemical Co.	D.E.R., D.E.N., D.E.H., Tactix, Quatrex
CIBA-GEIGY Corp.	Araldite
Dainippon Ink and Chemicals	Epotuf, Kelpoxy
Union Carbide Co.	Unox (cycloaliphatic diepoxides)

Health and Safety Factors

There have been many investigations of the toxicity of various classes of epoxy-containing materials (glycidyoxy compounds). The use and interpretation of the vast amount of data available has been obscured by two factors: (1) proper identification of the epoxy systems in question and (2) lack of meaningful classification of the epoxy materials. In general, the toxicity of many of the glycidyoxy derivatives is low but the diversity of compounds found within this group does not permit broad generalizations for the class.

Diglycidyl Ether of Bisphenol A. The liquid DGEBPA-based resins exhibit low acute toxicity with a single-dose oral LD₅₀ value in rats of >2000 mg kg (40). The potential for absorption of DGEBPA through the skin in acutely toxic amounts is low. LD₅₀ values of >800 mg kg for acute dermal toxicity have been obtained from studies using both the pure and commercial DGEBPA (41,42).

Some skin sensitization to low molecular-weight DGEBA resins (mol wt ~340) has been shown in animals and humans. Skin sensitization decreases with an increase in molecular weight but the presence of low molecular-weight fractions in the advanced resins may present a hazard to skin sensitization (43).

Inhalation toxicity does not present a hazard because of low vapor pressure. DGEBA-based resins have been reported to cause minimal eye irritation (44). Systemic toxicity has not been noted in experiments where DGEBA-based resins have been fed to laboratory animals. Mutagenic activity has not been shown in animals, but *in vitro* mutagenicity tests have yielded variable results (44).

Carcinogenicity of DGEBA or DGEBA-based resins, as measured by topical application, has not been shown by a majority of the studies (45). Advanced DGEBA resins exhibit low systemic toxicity either by dermal or oral routes and inhalation of these resins is unlikely because of low volatility. The acute oral LD₅₀ in rats has been reported to be >2000 mg/kg (46). Acute dermal studies show these materials have a low potential for absorption through the skin in acutely toxic amounts. No evidence of carcinogenicity has been found in animals or humans for advanced DGEBA resins (47,48).

Epoxy Phenol Novolak Resins. Acute oral studies indicate low toxicity for these resins (49). Eye studies indicate only minor irritation in animals (49). The EPN resins have shown weak skin-sensitizing potential in humans.

Applications

Protective Coatings. The largest single use of epoxy resins is in the protective coatings market where high chemical resistance and adhesion is important (see COATINGS). As a result of new ecology and energy requirements, increasing amounts of waterborne, high solids, and solventless systems (powder and liquid) are used. Cure mechanisms can be either one-package, heat-curing (baking) finishes, or room temperature-curing systems (including air-dried esters).

Ambient-cure systems are often based on lower molecular-weight solid epoxy resins cured with aliphatic polyamines or polyamides. Curing normally occurs at ambient temperatures with a working life (pot life) of 8–24 h, depending on the formulation. Epoxy–polyamine systems are typically used for maintenance coatings in oil refineries, petrochemical plants, and in many marine applications. Such coverings are applied by spray or brush. These are used widely where water immersion is encountered, particularly in marine applications (see COATINGS, MARINE).

The higher molecular-weight solid epoxy resins are used in formulations that usually consist of a resin, hardener, reinforcing filler, pigments, flow control agents, and other modifiers. In addition to using conventional hardeners in these formulations, epoxy resins can also be hardened with other resins, ie, acrylics or polyesters.

Ester-type air-dried coating systems provide better performance than alkyd coatings but lesser performance than straight epoxies (see ALKYD RESINS). These coatings are simple esterification products of the epoxy resin with a fatty acid, resin acid, or a tall oil acid (see CARBOXYLIC ACIDS; TALL OIL). They are available in short, medium, and long oil lengths similar to alkyds. Long oil length esters have limited chemical resistance, whereas the short oil length esters provide improved chemical resistance. The short oil length esters promote surface wetting similar to drying oils; therefore, they are often used when optimum surface preparation is not possible.

Heat-curing finishes generally are based on the higher molecular-weight solid epoxy resins cross-linked with phenol, melamine, or urea–formaldehyde condensation products (phenoplasts and aminoplasts). Normal epoxy–aminoplast weight ratios in resins are epoxy–urea, 70:30; epoxy–benzoguanamine, 70:30; epoxy–melamine, 80:20, 90:10. Increasing the epoxy–aminoplast ratio provides: (1) improved flexibility, toughness, adhesion, and gloss; (2) reduced gloss retention, solvent resistance, and hardness; and (3) higher temperature baking schedules. In conjunction with the need to reduce energy consumption, accelerators are used and faster-curing hardeners are being developed.

Epoxy–phenolic systems offer chemical resistance along with excellent mechanical properties. Resins of 4000–5000 mol wt are used in conjunction with a phenolic that has a low degree of condensation. Part of the methylol groups may be etherified with butanol. Generally, the ratio of epoxy–phenolic ranges from 70:30 to 80:20, by weight, depending on the properties desired. Acid catalysts such as phosphoric acid or the morpholine salt of *p*-toluenesulfonic acid are often used to accelerate the cure. Such systems are used for chemically resistant coatings on process equipment, tank and drum linings, and in pipe lining.

Waterborne Coatings. These coatings utilize either liquid or solid epoxy resins that have been modified to allow their use with water. They are usually in the form of emulsions, suspensions, dispersions, or water-dilutable resins that can be heat- or RT-cured. They are applied by convenient methods such as roller-coating, dipping, spray, or electrodeposition.

Owing to governmental regulations, considerable research has been expended to develop systems suitable for substitution of solvent-based systems, particularly for automobile and container applications. In the switch from solvent-based to waterborne systems, epoxies are successfully bridging the gap largely by adaptation of conventional resins.

Significant advances in waterborne coatings have been made by PPG Industries utilizing epoxies as co-resins. These coatings are used in cathodic electrodeposited systems, widely accepted for automobile primers. Many patents have been issued for this important technology (50,51).

A waterborne system for container coatings was developed based on a graft copolymerization of an advanced epoxy resin and an acrylic (52). The acrylic–vinyl monomers are grafted onto preformed epoxy resins in the presence of a free-radical initiator; grafting occurs mainly at the methylene group of the aliphatic backbone on the epoxy resin. The polymeric product is a mixture of methacrylic acid–styrene copolymer, solid epoxy resin, and graft copolymer of the unsaturated monomers onto the epoxy resin backbone. It is dispersible in water upon neutralization with an amine before cure with an amino–formaldehyde resin.

Solventless Coatings. Ecological concerns have led to increasing uses of these materials. The advantages include high buildup in a single application, minimization of surface defects owing to the absence of solvents, excellent heat and chemical resistance, and lower overall application costs. Disadvantages include poor impact resistance and flexibility, short pot life, and increased sensitivity to humidity.

Attempts to overcome the short pot life have focused on development of special equipment in which the components can be mixed in the proper proportion. Another approach has been to develop curing agents that give long pot life compared to conventional systems.

High Solids Coatings. High solids coatings resemble the technology of solvent-free coatings but the compositions contain ca 70% by volume of solid resin and are modified by reactive diluents, low viscosity multifunctional resins, or backbone structures other than the bisphenol A moiety.

Powder Coatings. Epoxy-based powder coatings exhibit useful properties such as excellent adhesion, abrasion resistance, hardness, and chemical resistance. The application possibilities are diverse, including refrigerator liners, oil filters, hospital equipment, primers, shelving, automobile springs, and fire extinguishers.

Interest in powders is primarily related to meeting increasingly stringent antipollution regulations. The development of highly reactive powder systems which cure at low energy (150°C) and the possibility of economical thin films (30–40 μm) have made powder coatings competitive with waterborne and high solids systems. Powder coatings can be applied by fluidized-bed (thick films, 50–150 μm) or electrostatic spray (thin films, 30–40 μm).

In powder coatings, epoxies are continuing to grow at rates exceeding other technology segments mainly because of the 100% solids feature, improved coverage, and recyclability of material. Pipeline projects, important because of worldwide energy problems, are significant users of epoxy resins. The maintenance coating segment, although not spectacular in its overall growth, still provides a steady and solid base for epoxies. The value of improved service life is being increasingly accepted even at the somewhat higher material cost of epoxy systems.

To improve the weatherability of epoxies, which normally chalk and yellow, epoxy–polyester alloys or hybrids are used. These powders with improved overbake resistance cure at temperatures as low as 130°C. They have film flexibility similar to epoxy resins, but their hardness is slightly decreased. Corrosion resistance is equivalent to epoxy powders in most cases, although solvent and alkali resistance is inferior.

Another approach to weatherable systems is the use of triglycidyl isocyanurate (6) as a curing agent for carboxyl-containing oil-free polyesters. The trifunctionality is responsible for its high reactivity with carboxyl-containing oil-free polyesters. These types of powder coatings are used in the protection of metal window frames, as well as for exterior siding.

Structural Composites. Because of their excellent adhesion, good mechanical, humidity, and chemical-resistance properties, epoxy resins are

used extensively in combination with glass, graphite, boron, and Kevlar fibers in the manufacture of high performance composites (see COMPOSITE MATERIALS).

An important application is for filament-wound glass-reinforced pipe used in oil fields, chemical plants, water distribution, and as electrical conduits. Low viscosity liquid systems having good mechanical properties (elongation at break) when cured are preferred. These are usually cured with liquid anhydride or aromatic-amine hardeners. Similar systems are used for filament-winding pressure bottles and rocket motor casings.

In the aerospace industry, the use of graphite fiber-reinforced composites is growing rapidly because of high strength-to-weight ratios (see ABLATIVE MATERIALS). High performance polyfunctional resins, such as the tetraglycidyl derivative of methylenedianiline in combination with diaminodiphenylsulfone, are extensively used to provide good elevated temperature properties and humidity resistance. Handling characteristics are well suited to the autoclave molding technique primarily used in the manufacture of such components.

Improved versions of the high performance resin systems continue to be developed (53). Toughening of epoxies has emerged as an important area for investigation using both rubber and thermoplastics (54–56).

Electrical Laminates. A significant use for epoxy resins is in the manufacture of copper-clad epoxy-glass printed circuit boards. Systems are available that meet the National Electrical Manufacturers Association (NEMA), G10, G11, FR3, FR4, FR5 specifications. Currently the majority of boards are manufactured to the flame-retardant FR4 specification. The flame retardance is achieved by the use of a solid epoxy resin based on tetrabromobisphenol A (see FLAME RETARDANTS). In combination with dicyandiamide as the hardener, stable impregnation varnishes can be formulated. The impregnated glass fabric can be dried and "B-staged" in a controlled manner by treating in a heated tunnel or tower. The B-staged fabric (prepreg) is cut into sheets and laid up into laminates with a copper foil top and bottom, and pressed at high pressure and elevated temperature to form a well-consolidated fully cured board. Identical procedures are used for making structural and electrical laminates usually based on nonflame-retardant resins.

Photoactive polymers containing epoxy groups are used in the manufacture of high performance printed wiring boards. A polymer with a uv-sensitive group allowing photocross-linking and an epoxy group for thermal cross-linking has been used as a photopolymerizable solder mask and in photoresists (57).

Electrical, Electronics, and Structural Applications.

Molding Compounds. Thermoset epoxy molding compounds are multicomponent mixtures usually based on solid epoxy resins, hardeners, and various fillers and reinforcements. The exact nature of the compounds is usually dictated by the application and the molding procedure to be used, ie, transfer, compression, or injection molding. Because of the use of narrow gates in transfer and injection molding processes, the use of long-fiber reinforcement (>3 mm) is usually avoided because of the danger of excessive orientation of the fibers. In compression-molding procedures, this limitation does not apply and long-fiber reinforced molding compounds are frequently used. Because of the good adhesive properties and low shrinkage on cure of epoxy resins, internal release agents are frequently incorporated into the molding compound formulation. Care must be taken in designing the mold to allow sufficient draft to facilitate easy ejection of the part on demolding.

The main use of epoxy molding compounds is for the encapsulation via transfer molding of solid-state devices such as diodes, transistors, and integrated circuits (see EMBEDDING). The molding compounds are formulated to meet stringent requirements regarding flow, reactivity, electrical properties, humidity, and thermal resistance. Extensive testing and experience have shown that the best characteristics are achieved using a solid, multifunctional, epoxidized cresol–novolak resin, and a phenol or cresol novolak as a hardener.

Low ionic impurity levels are imperative. In order to reduce the coefficient of thermal expansion of the final molding, and hence minimize stresses on the encapsulated silicon chip, the highest possible filler loading is desired. This has to be balanced against the need to maintain as low a melt viscosity as possible to minimize the possibility of damage to the device during the encapsulation process.

Epoxy molding compounds based on solid bisphenol A resins, frequently upgraded with solid epoxy phenol or cresol novolaks, and cured with novolaks or occasionally with solid anhydrides or aromatic amines, are used for encapsulating passive electronic components, ie, coils, capacitors, transformers, ignition coils, etc. In such applications small amounts of short glass fiber in addition to particulate filler may be used to increase impact strength (see FILLERS; GLASS). Similar compounds frequently are used for molding high voltage electrical bushings and connectors used in pad-mounted distribution transformers (see ELECTRICAL CONNECTORS).

Compression molding is mainly used in the manufacture of large, long-glass-fiber (up to 25 mm) reinforced parts. Chopped prepreg molding compounds, in which glass roving has been impregnated with a solid resin-hardener formulation and chopped to give strands of the desired length (usually 12–25 mm), and epoxy sheet-molding compounds are mainly used. These are characterized by excellent mechanical strength, rigidity, and good thermal or chemical resistance properties, or both. Components such as pump housings, impellers, pipe fittings, and valves are compression molded.

Casting and Encapsulation. Cast components are made by filling a mold cavity with a liquid epoxy system and then cross-linking the resin to give an infusible solid part. The resin system used may be formulated to cure either at RT or at elevated temperatures. A problem with casting large parts is that during cure, the excess heat of reaction must be dissipated to avoid large locked-in thermal stresses in the casting. This is achieved by skillful formulation but results in relatively long mold cycles and poor productivity.

Nevertheless, casting processes have found extensive uses where the good electrical and mechanical properties of epoxy resins offer many design advantages. The most commonly used systems are based on bisphenol A epoxy resins cured with anhydride hardeners and usually heavily (60–65 wt %) filled with silica fillers. Components using from 50–100 kg of mix are cast on a routine basis. Typical components manufactured by casting techniques are post insulators, busbar supports, switchgear components, instrument transformers, distribution transformers, high and low voltage bushings, encapsulated coils, etc. Poor wet arc-track resistance and uv resistance limit bisphenol A-based resins to indoor applications. For outdoor applications, the use of cycloaliphatic resins cured with a nonaromatic anhydride hardener, eg, hexahydrophthalic anhydride, is mandatory. Such systems have increasing use for outdoor insulators, switchgear components, and instrument transformers.

A casting process, pressure gelation (58), was developed that overcomes problems associated with the high heat of reaction of an epoxy resin. It allows large parts to be gelled rapidly, reducing the mold cycle from hours to minutes. The increased productivity is responsible for growing worldwide use in the electrical industry (see EMBEDDING).

In addition to electrical uses, epoxy casting resins are utilized in the manufacture of tools, ie, contact and match molds, stretch blocks, vacuum-forming tools, and foundry patterns, as well as bench tops and kitchen sinks. Systems consist of a gel-coat formulation designed to form a thin coating over the pattern which provides a perfect reproduction of the pattern detail. This is backed by a heavily filled epoxy system which also incorporates fiber reinforcements to give the tool its strength. For moderate temperature service, a liquid bisphenol A epoxy resin with an aliphatic amine is used. For higher temperature service, a modified system based on an epoxy phenol novolak and an aromatic diamine hardener may be used.

Adhesives. Because of excellent adhesion to many substrates, epoxy resins are extensively used for high performance adhesives. These can be categorized into high temperature curing systems (solids and liquids) and room temperature curing systems (liquids).

For the high temperature systems, dicyandiamide and aromatic-amine hardeners are frequently used. For room temperature systems, polyamide aminoamide, and aliphatic-amine hardeners are preferred.

Depending on the characteristics and performance requirements, adhesives systems are frequently modified with diluents (reactive and nonreactive) and polyfunctional high performance resins, as well as with fillers of various types.

In the adhesives market growth has been slower than in other segments. Increasing competition from other materials, most notably polyurethanes and acrylics, is affecting the market (see ACRYLIC ESTER POLYMERS; URETHANE POLYMERS).

BIBLIOGRAPHY

"Epoxy Resins" in *ECT* 1st ed., Suppl. Vol, pp. 312–329, by R. A. Coderre, Shell Chemical Corp.; in *ECT* 2nd ed., Vol. 8, pp. 294–312, by J. R. Weschler, Ciba Products Co.; in *ECT* 3rd ed., Vol. 9, pp. 267–290, by S. Sherman, J. Gannon, and G. Buchi, CIBA-GEIGY Corp., and W. R. Howell, Dow Chemicals U.S.A.

1. U.S. Pat. **2,324,483** (July 20, 1943), P. Castan (to CIBA-GEIGY Corp.).

2. U.S. Pat. **2,456,408** (Dec. 14, 1948), S. O. Greenlee (to DeVoe & Raynolds).

3. *Epoxy Cresol Novolac Resins*, technical data bulletin, CIBA-GEIGY Corp., Ardsley, N.Y., 1981.

4. *Araldite MT 0163*, technical data bulletin, CIBA-GEIGY Corp., Ardsley, N.Y., 1984; *EPON 1031*, technical data bulletin, Shell Chemical Co., Houston, Tex., 1990.

5. W. Hofmann and W. Fisch, *FATIEPEC Congr. Rev.*, **243** (1962).

6. *Araldite MY 720*, technical data bulletin, CIBA-GEIGY Corp., Ardsley, N.Y., 1979.

7. U.S. Pat. **2,951,825** (Sept. 6, 1960), N. H. Reinking and co-workers (to Union Carbide Co.); *Araldite MY 0500*, technical data bulletin, CIBA-GEIGY Corp., Ardsley, N.Y., 1979.

8. *Araldite PT 810*, technical data bulletin, CIBA-GEIGY Corp., Ardsley, N.Y., 1981.

9. W. G. Potter, *Epoxide Resins*, Springer-Verlag, New York, 1970.

10. U.S. Pat. **3,477,990** (Nov. 11, 1969), M. F. Dante and H. L. Parry (to Shell); **3,978,027** (Nov. 11, 1969), C. D. Marshall (to Shell); **4,048,141** (Sept. 13, 1977), G. A. Doorakian and co-workers (to The Dow Chemical Company).

11. J. J. Harris and S. C. Temin, *J. Appl. Polym. Sci.* **10**, 523 (1966).

12. A. J. Landua, *Am. Chem. Soc. Div. Org. Coat. Plast. Chem. Pap.* (42), (1964).

13. L. H. Lee and co-workers, *J. Polym. Sci. Part A* **3**, 2955 (1965).

14. R. J. Arnold, *Mod. Plast.* **41**, 149 (1964).

15. R. J. Morgan and E. T. Mones, *J. Appl. Polym. Sci.* **33**, 999 (1987).

16. L. Schechter and J. Wynstra, *Ind. Eng. Chem.* **48**, 86 (1956).

17. E. S. Narracott, *Br. Plast.* **28**, 120 (1953).

18. L. Schechter, J. Wynstra, and R. P. Kurjy, *Ind. Eng. Chem.* **48**, 94 (1956).

19. L. J. Gaugh and L. T. Smith, *J. Oil Colour Chem. Assoc.* **43**, 409 (1960).

20. R. E. Parker and N. S. Isaacs, *Chem. Rev.* **59**, 737 (1959).

21. N. B. Chapman, R. E. Parker, and N. S. Isaacs, *J. Chem. Soc.* **2**, 1925 (1959).

22. K. Dusek, M. Ilavsky, and S. Lunak, *J. Polym. Sci. Polym. Symp.* **53**, 29 (1975).

23. J. Charlesworth, *J. Polym. Sci.* **18**, 621 (1980).

24. K. Dusek, *Makromol. Chem. Macromol. Symp.* **7**, 37–53 (1987).

25. K. Dusek, *Adv. Polym. Sci.* **78**, 1–59 (1986).

26. M. Gaschke and B. Dreher, *J. Coatings Technol.* **48**(617), 46 (1976).

27. W. Fisch and W. Hofmann, *J. Polym. Sci.* **12**, 497 (1954).

28. W. Fisch, W. Hofmann, and J. Koskikallio, *J. Appl. Chem. London* **6**, 429 (1956).

29. R. F. Fischer, *J. Polym. Sci.* **44**, 155 (1960).

30. G. C. Stevens, *J. Appl. Polym. Sci.* **26**, 4259 (1981).

31. L. Mateska and co-workers, *J. Polym. Sci. Polym. Chem. Ed.* **21**, 2873 (1983).

32. S. A. Zahir, *Adv. Org. Coat. Sci. Technol.* **4**, 83 (1982).

33. A. Pfitzmann, K. Schlothauer, and M. Fedtke, *Polym. Bull.* **27**(1), 59–66 (1991).

34. A. Farkas and P. F. Strohm, *J. Appl. Polym. Sci.* **12**, 159 (1968).

35. F. Ricciardi, W. Romanchik, and M. Joullie, *J. Polym. Sci. Polym. Chem. Ed.* **21**, 1475 (1983).

36. H. J. L. Bressers and L. Goumans, in B. Sedlacek and J. Kahovec, eds., *Cross-Linked Epoxies*, de Gryter, Berlin, 1987, pp. 223–230.

37. M. Haley with R. Mulach and Y. Sakuma, "Epoxy Resins," in *Chemical Economics Handbook*, Stanford Research Institute International, Menlo Park, Calif., 1991.

38. *Thermosetting and Thermoplastic Resins*, year-end monthly or annual statistical reports, the Committee on Resin Statistics of the Society of the Plastics Industry, Inc., Washington, D.C., through 1991.

39. *Synthetic Organic Chemicals, U.S. Production and Sales*, U.S. International Trade Commission, 1991.

40. D. D. Lockwood and H. W. Taylor, *Acute Toxicological Properties and Industrial Handling Hazards of D.E.R. 361 Epoxy Resin*, The Dow Chemical Company, Midland, Mich., 1982.

41. R. W. Hend, D. G. Clark, and A. D. Coombs, *Toxicity of Purified Diglycidyl Ether of Bisphenol A: Results of Preliminary Studies*, Shell Report TLGR.0010.77, Shell Chemical Corp., Houston, Tex., 1977.

42. *Ibid.*, *Toxicity of EPIKOTE 828: Results of Preliminary Short-Term Studies*, Shell Report TLGR.0009.77. Shell Chemical Corp., Houston, Tex., 1977.

43. S. Fregert, in H. I. Maibach and G. A. Gellin, eds., *Occupational and Industrial Dermatology*, Year Book Medical Publishers, Chicago, 1982, pp. 185–288.

44. C. Hine and co-workers, in G. D. Clayton and F. E. Clayton, eds., *Patty's Industrial Hygiene and Toxicology*, Vol. 2A, John Wiley & Sons, Inc., New York, 1981, pp. 2141–2258.

45. J. M. Holland and co-workers, *Chronic Dermal Toxicity of Epoxy Resins I. Skin Carcinogenic Potency and General Toxicity*, report ORNL-5762, prepared for the U.S. DOE by Oak Ridge National Laboratory, Oak Ridge, Tenn., 1981.

46. T. G. Pullin, *Integrated Mutagenicity Testing Program on Several Epoxy Compounds*, The Dow Chemical Company, 1977.

47. S. Fregert and A. Thorgeirsson, *Cont. Derm.* **3**, 301–303 (1977).

48. D. C. Mensik and D. D. Lockwood, *DER 662 UH Epoxy Resin: Dermal Sensitization in Hartley Guinea Pigs*, The Dow Chemical Company, 1987.

49. D. D. Lockwood and V. Barrego, *Acute Toxicological Properties and Industrial Handling Hazards of Experimental Resin SD-7855.05*, The Dow Chemical Company, 1979.

50. U.S. Pat. **3,947,339** (Mar. 30, 1978), R. D. Jerabek and co-workers (to PPG Industries, Inc.).

51. U.S. Pat. **4,009,133** (Aug. 1, 1975), J. E. Jones (to PPG Industries, Inc.).

52. U.S. Pat. **4,212,781** (July 15, 1980), J. M. Evans and V. W. Ting (to SCM Corp.).

53. M. R. Thoseby, B. Dobinson, and C. H. Bull, *Br. Polym. J.* **18**, 286 (1986).

54. W. D. Bascom and D. L. Hunston, *ACS Symp. Ser.* **222**, 135 (1989).

55. R. Drake, *ACE Div. Poly. Mats. Sci. Eng.* **63**, 802 (1990).

56. R. Mülhaupt, *Chimia* **44**, 43 (1990).

57. G. E. Green, B. P. Stark, and S. A. Zahir, *J. Macromol. Sci. Revs. Macromol. Chem.* **21**, 187 (1981–1982).

58. U.S. Pat. **3,777,000** (Dec. 4, 1973), E. Kusenbergh and co-workers (to CIBA-GEIGY Corp.).

General References

H. Lee and K. Neville, *Handbook of Epoxy Resins*, McGraw-Hill, Inc., New York 1967, Reprinted 1982.

W. G. Potter, *Epoxide Resins*, Springer-Verlag, New York, 1970.

The Epoxy Resin Formulators Training Manual, The Society of Plastics Industry, Inc., New York, 1984.

C. A. May, ed., *Epoxy Resins Chemistry and Technology*, 2nd ed., Marcel Dekker, Inc., New York, 1988.

L. V. McAdams and J. A. Gannon, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 6, Wiley-Interscience, New York, 1986, pp. 322–382.

Toxicology

M. O. Amdur, J. Doull, and C. D. Klaassen, eds., *Casarett and Doull's Toxicology*, 4th ed., Pergamon Press, New York, 1991.
M. Ellenhorn and D. G. Barceloux, *Medical Toxicology*, Elsevier, New York, 1988.
F. N. Marzulli and H. I. Maibach, *Dermatotoxicology*, Hemisphere Publishing Corp., Washington, D.C., 1991.
G. D. Clayton and F. Clayton, eds., *Patty's Industrial Hygiene and Toxicology*, 3rd rev. ed., John Wiley & Sons, Inc., New York, 1978.
W. M. Grant, *Toxicology of the Eye*, Charles C Thomas, Springfield, Ill., 1988.

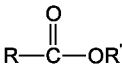
John Gannon
Consultant

ERBIUM.

See LANTHANIDES.

ESTERIFICATION

This article describes methods for the production of carboxylic esters:



For the properties of these compounds, see ESTERS, ORGANIC. For esters of inorganic acids, see the articles on nitric acid, phosphoric acids, sulfuric acid, etc.

Esters are most commonly prepared by the reaction of a carboxylic acid and an alcohol with the elimination of water. Esters are also formed by a number of other reactions utilizing acid anhydrides, acid chlorides, amides, nitriles, unsaturated hydrocarbons, ethers, aldehydes, ketones, alcohols, and esters (via ester interchange). Detailed reviews of esterification are given in References 1–9.

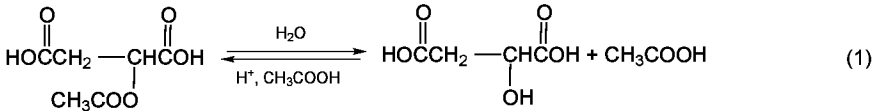
On the basis of bulk production (10), poly(ethylene terephthalate) manufacture is the most important ester producing process. This polymer is produced by either the direct esterification of terephthalic acid and ethylene glycol, or by the transesterification of dimethyl terephthalate with ethylene glycol. In 1990, poly(ethylene terephthalate) manufacture exceeded 3.47×10^6 t yr (see POLYESTERS). Dimethyl terephthalate is produced by the direct esterification of terephthalic acid and methanol.

Other large-volume esters are vinyl acetate [108-05-4] (VAM, 1.15×10^6 t yr), methyl methacrylate [80-62-6] (MMA, 0.54×10^6 t yr), and dioctyl phthalate [117-81-7] (DOP, 0.14×10^6 t yr). VAM (see VINYL POLYMERS) is produced for the most part by the vapor-phase oxidative acetoxylation of ethylene. MMA (see METHACRYLIC POLYMERS) and DOP (see PHTHALIC ACIDS) are produced by direct esterification techniques involving methacrylic acid and phthalic anhydride, respectively.

The acetates of most alcohols are also commercially available and have diverse uses. Because of their high solvent power, ethyl, isopropyl, butyl, isobutyl, amyl, and isoamyl acetates are used in cellulose nitrate and other lacquer-type coatings (see CELLULOSE, ESTERS). Butyl and hexyl acetates are excellent solvents for polyurethane coating systems (see COATINGS; URETHANE POLYMERS). Ethyl, isobutyl, amyl, and isoamyl acetates are frequently used as components in flavoring (see FLAVORS AND SPICES), and isopropyl, benzyl, octyl, geranyl, linalyl, and methyl acetates are important additives in perfumes (qv).

Reactions Between Organic Acids and Alcohols

In the esterification of organic acids with alcohols, it has been shown that in most cases under acid catalysis, the union is between acyl and alkoxy groups. Acid hydrolysis of acetoxy succinic acid gives malic acid with retention of configuration at the asymmetric carbon atom (11):



n-Amyl alcohol produced by basic hydrolysis of *n*-amyl acetate with ¹⁸O-enriched water does not contain ¹⁸O (12).

Effect of Structure. The rate at which different alcohols and acids are esterified as well as the extent of the equilibrium reaction are dependent on the structure of the molecule and types of functional substituents of the alcohols and acids. Specific data on rates of reaction, mechanisms, and extent of reaction are discussed in the following. More details concerning structural effects are given in References 6, 13–15.

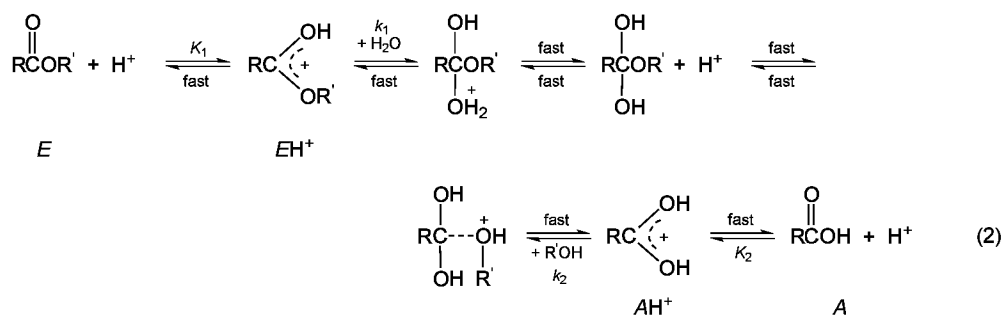
In making acetate esters, the primary alcohols are esterified most rapidly and completely, ie, methanol gives the highest yield and the most rapid reaction. Ethyl, *n*-propyl, and *n*-butyl alcohols react with about equal velocities and conversions. Under the same conditions, the secondary alcohols react much more slowly and afford lower conversions to ester products; however, wide variations are observed among the different members of this series. The tertiary alcohols react slowly, and the conversions are generally low (1–10% conversion at equilibrium). With isobutyl alcohol at 155°C, acids containing a straight-chain (acetic, propionic, and butyric) and phenylacetic and β-phenylpropionic acids are esterified readily. Formic acid has the highest initial rate of reaction. The introduction of a branched chain in the acid decreases the rate of esterification, and two branches cause a still greater retarding effect. However, the conversions to ester products from these substituted acids is higher than for the normal straight-chain acids. Similarly, aromatic acids, benzoic and *p*-toluic, react slowly but have high equilibrium conversions.

The introduction of a nitrile group on an aliphatic acid has a pronounced inhibiting effect on the rate of esterification. With the chloroacetic acids, the velocity decreases with increased chlorination. Double bonds also have a retarding influence on the rate of esterification. Tests on substituted acrylic acids have shown that α,β-unsaturated acids are esterified much less easily than the saturated analogues. A triple bond in the α,β position has about the same effect as a double bond. A β,γ-double bond has less of a retarding action. If the double bond is sufficiently removed, as in erucic and brassic acids (see CARBOXYLIC ACIDS), no effect is noted. Conjugated double bonds, when one is in the α,β-position, afford a great retarding effect. Cis-substituted unsaturated acids esterify more slowly than the trans isomers.

In the preparation of ethyl esters using anhydrous ethyl alcohol and hydrogen chloride catalyst, the rate of esterification of straight-chain fatty acids from propionic through stearic is substantially constant: branching of the fatty acid chain causes retardation. In the saturated dibasic acids, the rate of esterification is a maximum at glutaric acid. The ease of esterification of the cycloparaffin monocarboxylic acids increases in the order C3, C7, C6, C5, and C4 rings; with the exception of cyclopropanecarboxylic acid, these are esterified more rapidly than the corresponding open-chain acids.

Substitutions that displace electrons toward the carboxyl group of aromatic acids diminish the rate of the reaction (16). The substitution of fluoromethoxy or ethoxy groups in the ortho position has an accelerating action, whereas iodo, bromo, nitro, or methyl groups produce retardation. The influence of groups in the meta and para positions is not nearly so marked (17).

Kinetic Considerations. Extensive kinetic and mechanistic studies have been made on the esterification of carboxylic acids since Berthelot and Saint-Gilles first studied the esterification of acetic acid (18). Although ester hydrolysis is catalyzed by both hydrogen and hydroxide ions (19,20), a base-catalyzed esterification is not known. A number of mechanisms for acid- and base-catalyzed esterification have been proposed (4). One possible mechanism for the bimolecular acid-catalyzed ester hydrolysis and esterification is shown in equation 2 (6).



This mechanism leads to the rate equation (eq. 3) for hydrolysis and to an analogous expression for the esterification (13):

$$\frac{d[E]}{dt} = \frac{k_1 K_1 [E] [H_2O] [H^+]}{1 + \alpha} - \frac{k_2 K_2 [A] [R'OH] [H^+]}{1 + 1/\alpha} \quad (3)$$

In this expression, α depends on those rate coefficients in the above mechanism whose values are assumed to be high. Other mechanisms for the acid hydrolysis and esterification differ mainly with respect to the number of participating water molecules and possible intermediates (21–23).

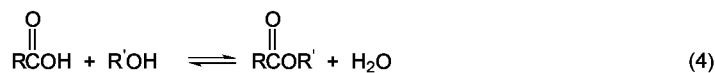
Applications of kinetic principles to industrial reactions are often useful. Initial kinetic studies of the esterification reaction are usually conducted on a small scale in a well stirred batch reactor (24). In many cases, results from batch studies can be used in the evaluation of the esterification reaction in a continuous operating configuration. Generally, the rate of esterification with acid catalyst is proportional to the acid or hydrogen ion concentration as well as the concentration of the alcohols and organic acid. The effect of temperature on the reaction rate is given by the well known Arrhenius equation. These factors are interrelated, and may be used to predict optimum operational conditions for the production of a given ester if the necessary data are available, i.e., the order of the reaction under the conditions to be used, a mathematical relation describing the yield with time, and an empirical equation relating the reaction rate constant with temperature, catalyst concentration, and proportions of reactants.

With these kinetic data and a knowledge of the reactor configuration, the development of a computer simulation model of the esterification reaction is invaluable for optimizing esterification reaction operation (25–28). However, all esterification reactions do not necessarily permit straightforward mathematical treatment. In a study of the esterification of 2,3-butanediol and acetic acid using sulfuric acid catalyst, it was found that the reaction occurs through two pairs of consecutive reversible reactions of approximately equal speeds. These reactions do not conform to any simple first-, second-, or third-order equation, even in the early stages (29).

In a study of the kinetics of the reaction of 1-butanol with acetic acid at 0–120°C, an empirical equation was developed that permits estimation of the value of the rate constant with a deviation of 15.3% from the molar ratio of reactants, catalyst concentration, and temperature (30). This study was conducted using sulfuric acid as catalyst with a mole ratio of 1-butanol to acetic acid of 3:19.6, and a catalyst concentration of 0–0.14 wt %.

Similar studies have been performed on the formation of mono *n*-butyl phthalate at 80–150°C with sulfuric acid catalyst (31). The reaction of phthalic anhydride with mono *n*-butyl phthalate to afford di *n*-butyl phthalate is complete in 10 min at 100°C with 1 wt % catalyst.

Equilibrium Constants. The reaction between an organic acid and an alcohol to produce an ester and water is expressed in equation 4:



This was first demonstrated in 1862 by Berthelot and Saint-Gilles (32), who found that when equivalent quantities of ethyl alcohol and acetic acid were allowed to react, the esterification stopped when two-thirds of the acid had reacted. Similarly, when equal molar proportions of ethyl acetate and water were heated together, hydrolysis of the ester stopped when about one-third of the ester was hydrolyzed. By varying the molar ratios of alcohol to acid, yields of ester >66% were obtained by displacement of the equilibrium. The results of these tests were in accordance with the mass action law shown in equation 5.

$$K \quad [ester] [water] \quad [acid] [alcohol] \quad (5)$$

However, in many cases the equilibrium constant is affected by the proportion of reactants (7,33,34). The temperature as well as the presence of salts may also affect the value of the equilibrium constant (35,36).

The effect of water on the equilibrium constant for the reaction of 1 mol of ethanol, 1 mol of acetic acid, and 23 moles of water has been investigated. This mixture has an equilibrium constant of 3.56, compared with 3.79 for the reaction with anhydrous materials (7,37).

Theoretical yields of ester obtainable with proportions of reactants are shown in Figure 1 for four values of the equilibrium constant. Thus when K equals 10 (esters of *p*-toluic acid with primary alcohols), with equivalent amounts of acid and alcohol, a yield of about 76% may be expected.

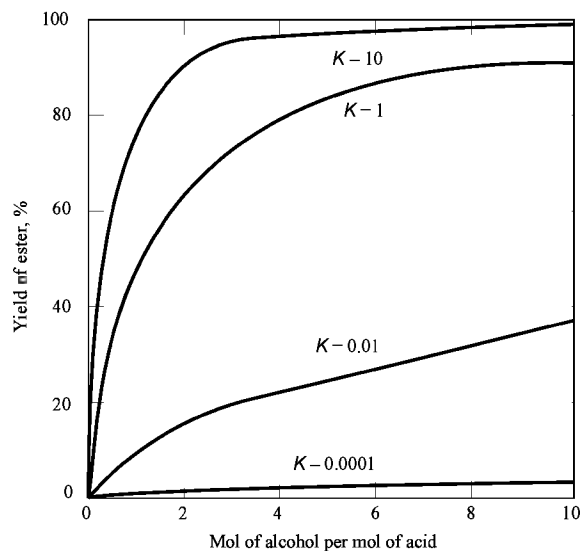


Fig. 1. Theoretical yields of ester obtainable with varying proportions of reactants for different values of equilibrium constant.

In general, esters having equilibrium constants below unity are not prepared by direct interaction of alcohol and acid; in these cases, the acid anhydrides or acid chlorides are used, since the equilibrium favors the ester product.

Completion of Esterification. Because the esterification of an alcohol and an organic acid involves a reversible equilibrium, these reactions usually do not go to completion. Conversions approaching 100% can often be achieved by removing one of the products formed, either the ester or the water, provided the esterification reaction is equilibrium limited and not rate limited. A variety of distillation methods can be applied to afford ester and water product removal from the esterification reaction (see DISTILLATION). Other methods such as reactive extraction and reverse osmosis can be used to remove the esterification products to maximize the reaction conversion (38). In general, esterifications are divided into three broad classes, depending on the volatility of the esters:

- (1) Esters of high volatility, such as methyl formate, methyl acetate, and ethyl formate, have lower boiling points than those of the corresponding alcohols, and therefore can be readily removed from the reaction mixture by distillation.
- (2) Esters of medium volatility are capable of removing the water formed by distillation. Examples are propyl, butyl, and amyl formates, ethyl, propyl, butyl, and amyl acetates, and the methyl and ethyl esters of propionic, butyric, and valeric acids. In some cases, ternary azeotropic mixtures of alcohol, ester, and water are formed. This group is capable of further subdivision: with ethyl acetate, all of the ester is removed as a vapor mixture with alcohol and part of the water, while the balance of the water accumulates in the system. With butyl acetate, on the other hand, all of the water formed is removed overhead with part of the ester and alcohol, and the balance of the ester accumulates as a high boiler in the system.
- (3) Esters of low volatility are accesible via several types of esterification. In the case of esters of butyl and amyl alcohols, water is removed as a binary azeotropic mixture with the alcohol. To produce esters of the lower alcohols (methyl, ethyl, propyl), it may be necessary to add a hydrocarbon such as benzene or toluene to increase the amount of distilled water. With high boiling alcohols, ie, benzyl, furfuryl, and β-phenylethyl, an accessory azeotrope liquid is useful to eliminate the water by distillation.

Use of Azeotropes to Remove Water. With the aliphatic alcohols and esters of medium volatility, a variety of azeotropes is encountered on distillation (see DISTILLATION, AZEOTROPIC AND EXTRACTIVE). Removal of these azeotropes from the esterification reaction mixture drives the equilibrium in favor of the ester product (39).

Binary azeotropes may be formed between the alcohol and water, the alcohol and ester, and the ester and water. Ternary azeotropes involving the alcohol, ester, and water are also possible. In general, the ternary azeotropes have the lowest boiling points, but the differences between the boiling points of the various combinations in some instances are very small. The ester–water binaries have boiling points close to those of the ternary mixtures. An extremely efficient fractionating column is usually required to obtain a pure ternary azeotrope. Binary azeotropes of the alcohol and water may be utilized in the preparation of the higher boiling, nonvolatile esters for completion of the reaction (39). Almost all of the alcohols (up to C20-alcohols) except methanol form binary azeotropes with water. The azeotropes formed by water with ethyl, *n*-propyl, isopropyl, allyl, and *tert*-butyl alcohols are single phase, ie, on condensation of the vapor, the components are completely miscible. Other means to eliminate water are often necessary: extraction of the ester with a water-insoluble solvent, eg, benzene, cyclohexane, or carbon tetrachloride; reactive distillation; drying with potassium carbonate; or salting out. The higher alcohols form azeotropes that on condensation separate into two liquid phases; in such a case, the alcohol-rich phase can be separated by further distillation into azeotrope and pure alcohol, and the water-rich phase into azeotrope and water. Under certain conditions, entraining gases are used to facilitate the removal of water (40).

Use of Desiccants and Chemical Means to Remove Water. Another means to remove the water of esterification is calcium carbide supported in a thimble of a continuous extractor through which the condensed vapor from the esterification mixture is percolated (41) (see CARBIDES). A column of activated bauxite (Florit) mounted over the reaction vessel has been used to remove the water of reaction from the vapor by adsorption (42).

Catalysts. The choice of the proper catalyst for an esterification reaction is dependent on several factors (43–46). The most common catalysts used are strong mineral acids such as sulfuric and hydrochloric acids. Lewis acids such as boron trifluoride, tin and zinc salts, aluminum halides, and organo–titanates have been used. Cation-exchange resins and zeolites are often employed also.

In laboratory preparations, sulfuric acid and hydrochloric acid have classically been used as esterification catalysts. However, formation of alkyl chlorides or dehydration, isomerization, or polymerization side reactions may result. Sulfonic acids, such as benzenesulfonic acid, *p*-toluenesulfonic acid, or methanesulfonic acid, are widely used in plant operations because of their less corrosive nature. Phosphoric acid is sometimes employed, but it leads to rather slow reactions. Soluble or supported metal salts minimize side reactions but usually require higher temperatures than strong acids.

Acid-Regenerated Cation Exchangers. The use of acid-regenerated cation resin exchangers (see ION EXCHANGE) as catalysts for effecting esterification offers distinct advantages over conventional methods. Several types of cation-exchange resins can be used as solid catalysts for esterification (47,48). In general, the strongly acidic sulfonated resins comprised of copolymers of styrene, ethylvinylbenzene, and divinylbenzene are used most widely. With the continued improvement of ion-exchange resins, such as the macroporous sulfonated resins, esterification has become one of the most fertile areas for use of these solid catalysts. With low molecular weight acids and alcohols, in most cases the resin structure has minimal effect on the yield or kinetics of the esterification as long as the catalyst contains strongly acidic groups. The kinetics in batch and tubular reactors of the esterification of 1-butanol with acetic acid catalyzed by a macroporous sulfonated polystyrene exchange resin have been studied. The catalytic activity was dependent on the water content of the resin and the rate determining step is the surface reaction of the chemisorbed acid and adsorbed alcohol (49).

Despite the higher cost compared with ordinary catalysts, such as sulfuric or hydrochloric acid, the cation exchangers present several features that make their use economical. The ability to use these agents in a fixed-bed reactor operation makes them attractive for a continuous process (50,51). Cation-exchange catalysts can be used also in continuous stirred tank reactor (CSTR) operation.

The resin (Amberlite IR-116 and Amberlite IR-120B) catalyzed continuous esterification of butanol or 2-ethylhexanol with acrylic acid is a novel example. High conversion and selectivity to the acrylate ester are accomplished. The CSTR in this application has advantage over a fixed-bed configuration since water separation from the higher boiling ester product in the reactor is more rapid, and this leads to a higher conversion with fewer by-products. The type of exchange resin also affects the esterification. The degree of cross-linking, porosity, and surface area of the strongly acid cation-exchange resin in combination with the back-mixed reactor design are critical factors to minimize secondary by-product formation and polymeric fouling of the resin catalyst (46).

The esterification of *n*-butyl alcohol and oleic acid with a phenol–formaldehydesulfonic acid resin (similar to amberlite IR-100) is essentially second order after an initial slow period (52). The velocity constant is directly proportional to the surface area of the catalyst per unit weight of reactants.

A series of tests using Amberlite IR-12 (sulfonated polystyrene resin) to esterify diethylene glycol (DEG) using toluene as the entrainer for removal of water gave the results in Table 1 (53).

Table 1. Tests Using Amberlite IR-120 to Esterify Diethylene Glycol (DEG)

Acid	DEG, mol	mol acid	IR-120, g	100 g acid	Temp, °C	Reaction time,	Monoester, %	Diester, %
						h	conv	conv
lauric	1		7.5		140	18	24	71
lauric	4		15.5		130	10	71	21
lauric	6		7.5		140	18	86	11
lauric	12		15.0		132	18	100	
oleic	12		10.6		140	18	100	
stearic	12		8.9		150	18	100	
benzoic	2		24.6		140	4	75	

Recovery of dilute acetic acid is achieved by esterification with methanol using a sulfonated resin (Dowex 50w) in a packed distillation column (54). Pure methyl acetate is obtained. This reaction is second order in acetic acid, zero order in methanol, and partially diffusion controlled.

Batch Esterification

Ethyl Acetate. A typical plant configuration for production of ethyl acetate [141-78-6] as a low boiling overhead product relative to water is shown in Figure 2 (2). The esterification reactor is a cylindrical tank, or still pot, heated by a closed-coil steam pipe. The reactor is charged with acetic acid, 95° ethanol, and concentrated sulfuric acid. The temperature at the top of the fractionating column is maintained at ca 70°C to give a ternary azeotropic mixture of ca 83° ethyl acetate, 9° alcohol, and 8° water. The vapor is condensed, part of it is returned to the top plate of the column as reflux, and the remainder is drawn off to storage. The ternary azeotrope (production-grade ethyl acetate) is satisfactory for many commercial purposes, but for an alcohol-free and water-free ester, further purification is needed.

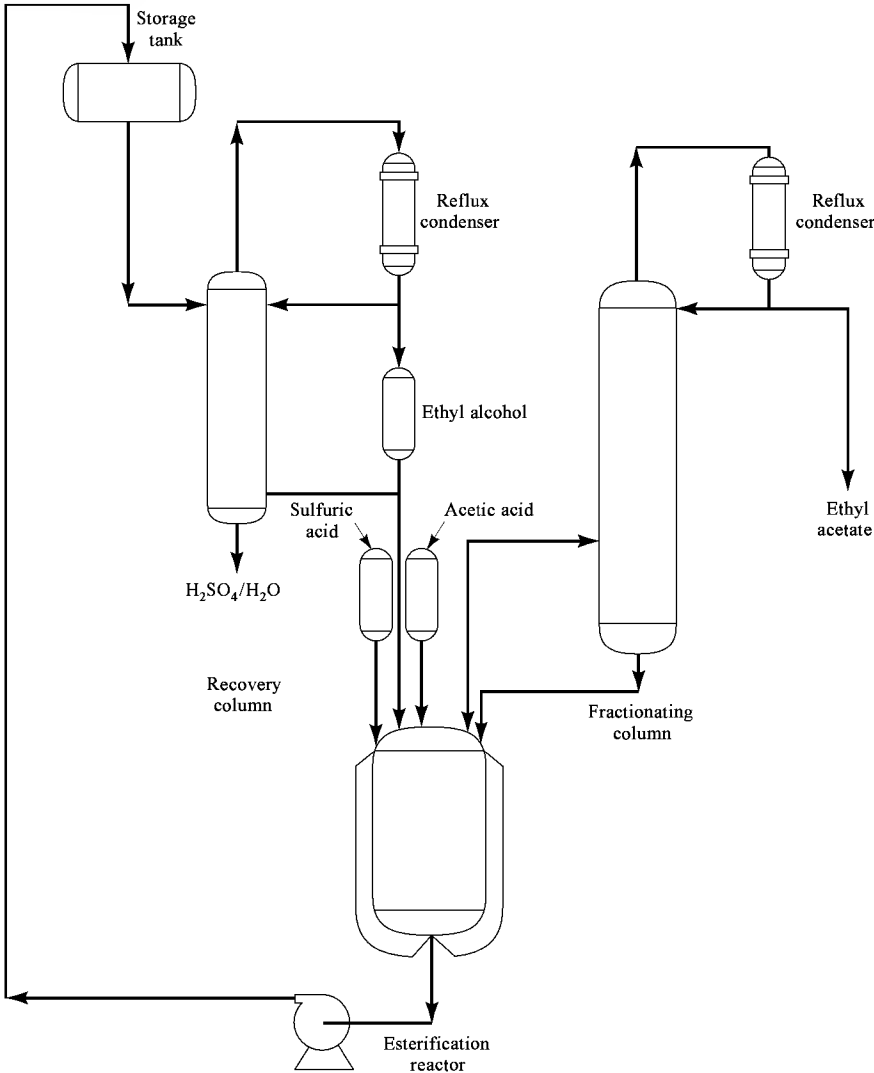


Fig. 2. Batch ethyl acetate process (2).

n-Butyl Acetate. Equipment used for the batch esterification to give butyl acetate [123-86-4] is shown in Figure 3. Glacial acetic acid is mixed with an excess of butyl alcohol and a small amount of concentrated sulfuric acid in the esterification reactor. The mixture is heated for several hours by means of a steam jacket to give esterification equilibrium. After the preliminary heating, slow rectification is permitted to remove the water already formed and thus increase the yield. The esterification is continued until no more water separates. At this point, the temperature at the top of the column rises, and the percentage of acetic acid in the distillate increases. It is necessary to neutralize the small amount of acid remaining in the esterification reactor before further distillation. A solution of sodium hydroxide is added to the esterification reactor, and the mixture is allowed to stand to form a water layer that is removed. The organic ester layer (upper layer) is then washed with water and distilled to obtain an overhead butyl acetate product of 75–85° purity; the remainder is butyl alcohol.

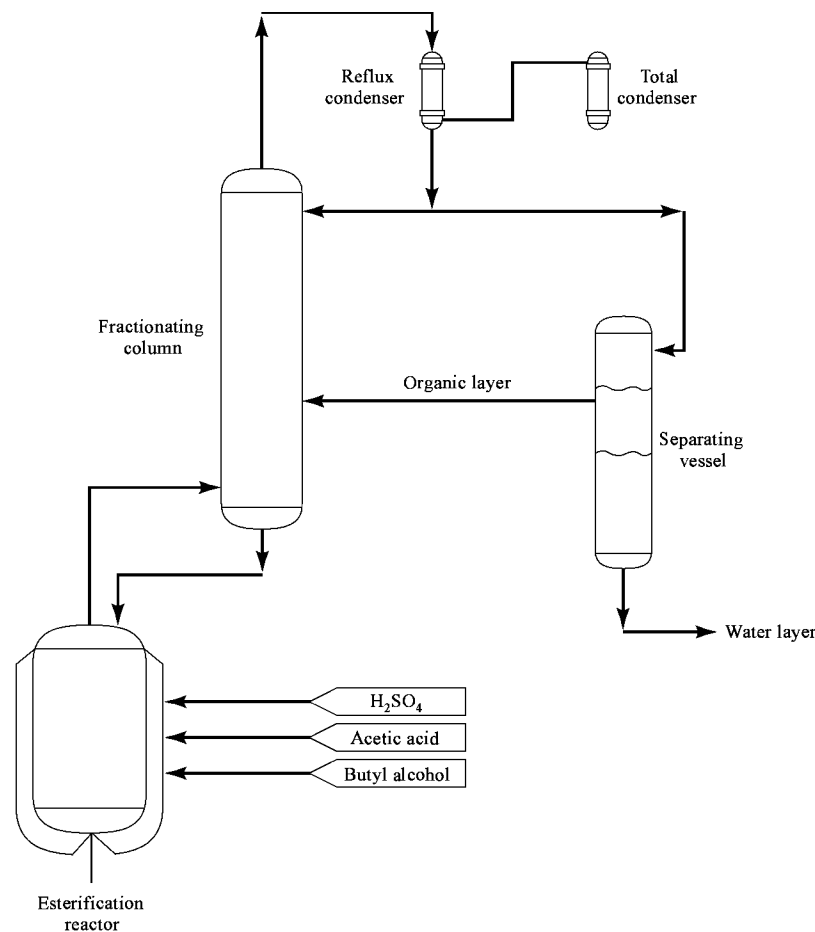


Fig. 3. Batch *n*-butyl acetate process (2).

Continuous Esterification

The law of mass action, the laws of kinetics, and the laws of distillation all operate simultaneously in a process of this type. Esterification can occur only when the concentrations of the acid and alcohol are in excess of equilibrium values; otherwise, hydrolysis must occur. The equations governing the rate of the reaction and the variation of the rate constant (as a function of such variables as temperature, catalyst strength, and proportion of reactants) describe the kinetics of the liquid-phase reaction. The usual distillation laws must be modified, since most esterifications are somewhat exothermic and reaction is occurring on each plate. Since these kinetic considerations are superimposed on distillation operations, each plate must be treated separately by successive calculations after the extent of conversion has been determined (see DISTILLATION).

Continuous esterification of acetic acid in an excess of *n*-butyl alcohol with sulfuric acid catalyst using a four-plate single bubblecap column with reboiler has been studied (55). The rate constant and the theoretical extent of reaction were calculated for each plate, based on plate composition and on the total incoming material to the plate. Good agreement with the analytical data was obtained.

A continuous distillation process has been studied for the production of high boiling esters from intermediate boiling polyhydric alcohols and low boiling monocarboxylic aliphatic or aromatic acids (56). The water of reaction and some of the organic acid were continuously removed from the base of the column.

Methyl Acetate. High purity methyl acetate [79-20-9] is required for the rhodium catalyzed carbonylation process to produce acetic anhydride (57). In the most recently developed commercial process for the manufacture of high purity methyl acetate, acetic acid functions both as a reactant and as an extractant in a countercurrent reactive distillation column (58,59), thereby alleviating the problem of azeotrope formation. This methyl acetate purification process obviates the use of additional vacuum or extractive distillation means to separate methyl acetate from its low boiling water and methanol azeotropes (60,61). As shown in Figure 4, this process uniquely demonstrates the use of reactive distillation as a means to produce essentially dry methyl acetate. The esterification reaction catalyzed by sulfuric acid occurs in the middle of the column. Acetic acid is fed to the top portion of the reactor section, and the methanol is fed to the lower portion of the reactor section. The countercurrent flow of acetic acid and methyl acetate with its azeotropes is used to remove water and by-products from methyl acetate. Below the acetic acid feed and above the reaction section, water and some methanol are extracted from methyl acetate using acetic acid. Acetic acid and methyl acetate are then separated above the acetic acid feed in the rectification portion of the column. High purity methyl acetate (at least 99.5 wt % methyl acetate) is isolated from the column overhead. The catalyst and impurities (primarily methyl propionate and isopropyl acetate) are removed from the reactor section by a sidedraw. Methanol in turn is stripped from the water in the lower portion of the column below the methanol feed. The impurities are further concentrated and removed from the process in two distillation columns with catalyst and acetic acid being recycled back to the reactive distillation column.

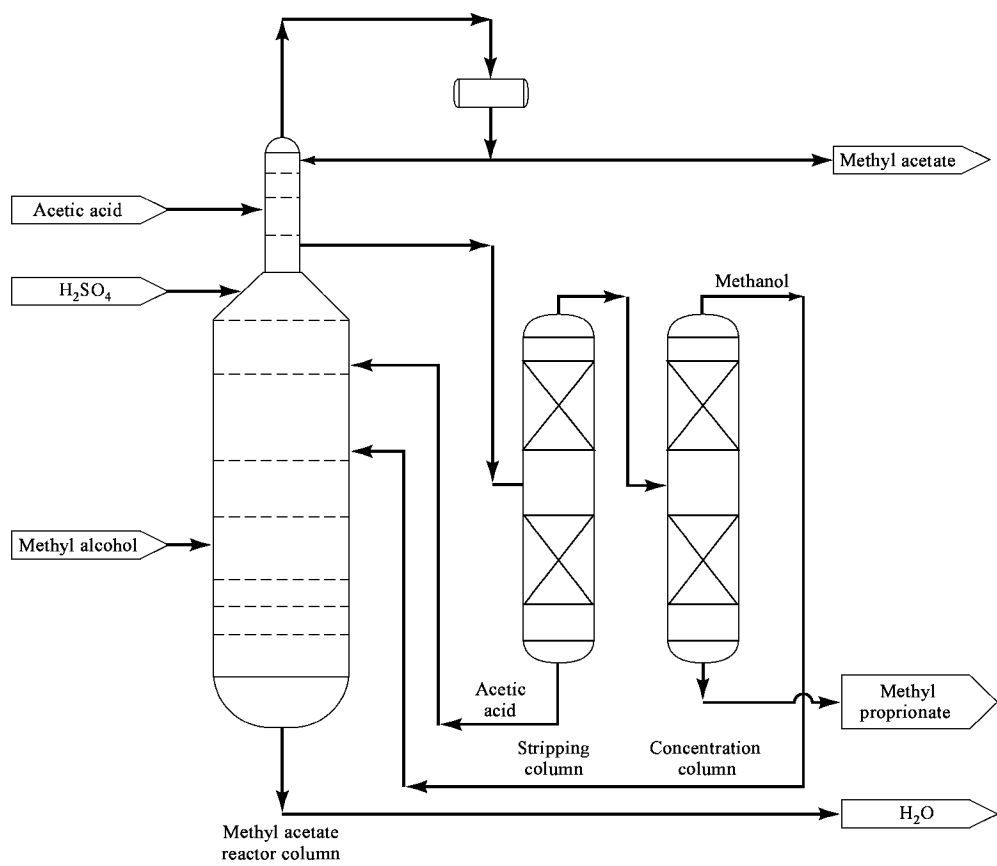


Fig. 4. Continuous methyl acetate process (59).

Ethyl Acetate. The production of ethyl acetate by continuous esterification is an excellent example of the use of azeotropic principles to obtain a high yield of ester (2). The acetic acid, concentrated sulfuric acid, and an excess of 95° ethyl alcohol are mixed in reaction tanks provided with agitators. After esterification equilibrium is reached in the mixture, it is pumped into a receiving tank and through a preheater into the upper section of a bubblecap plate column (Fig. 5). The temperature at the top of this column is maintained at ca 80°C and its vapor (alcohol with the ester formed and ca 10° water) is passed to a condenser. The first recovery column is operated with a top temperature of 70°C, producing a ternary azeotrope of 83° ester, 9° alcohol, and 8° water. The ternary mixture is fed to a static mixer where water is added in order to form two layers and allowed to separate in a decanter. The upper layer contains ca 93° ethyl acetate, 5° water, and 2° alcohol, and is sent to a second recovery or ester-drying column. The overhead from this column is 95–100° ethyl acetate which is sent to a cooler and then to a storage tank. This process also applies to methyl butyrate.

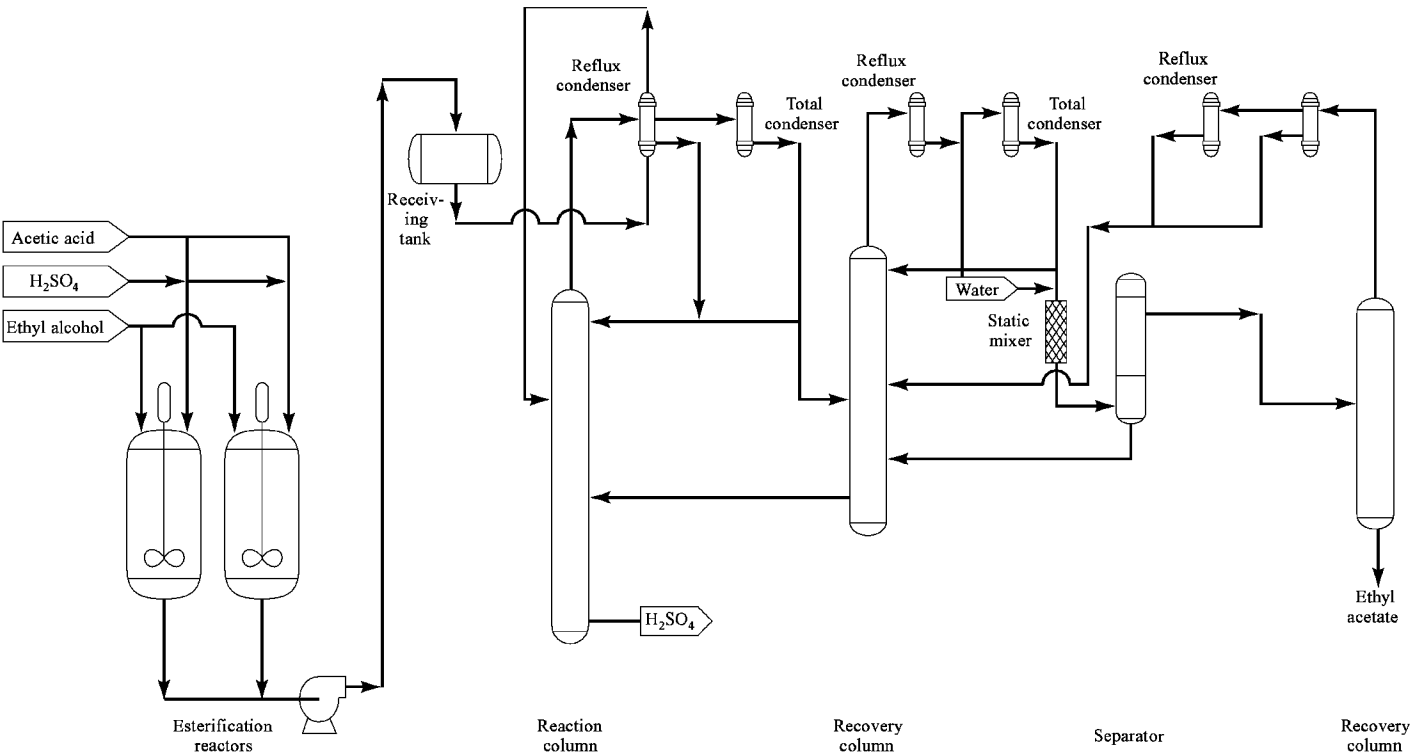


Fig. 5. Continuous ethyl acetate process (2).

Vapor-Phase Esterification

Catalytic esterification of alcohols and acids in the vapor phase has received attention because the conversions obtained are generally higher than in the corresponding liquid-phase reactions (7).

Physicochemical Considerations. The determination of the equilibrium constant K_C for the reaction $C_2H_5OH + CH_3COOH \rightleftharpoons C_2H_5OOCCH_3 + H_2O$ has been the subject of a number of investigations over the temperature range of 40–300°C (62). The

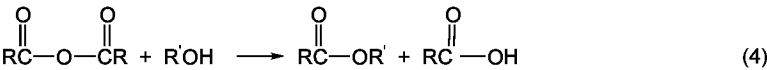
values of the equilibrium constant range from 6–559 (63) with 71–95° ester as the equilibrium concentration from an equimolar mixture of ethyl alcohol and acetic acid, depending on the technique used. A study of the reaction mechanism indicates that adsorption of acetic acid is the rate-controlling step; the molecularly adsorbed acetic acid then reacts with alcohol in the vapor phase. The rate of esterification of acetic acid and ethyl alcohol in equimolar quantities has been studied in a dynamic system using silica gel catalyst at 150–270°C (64).

Ethyl Acetate. Catalysts proposed for the vapor-phase production of ethyl acetate include silica gel, zirconium dioxide, activated charcoal, and potassium hydrogen sulfate. More recently, phosphoric-acid-treated coal (65) and calcium phosphate (66) catalysts have been described.

Other Esters. The esterification of acetic acid with various alcohols in the vapor phase has been studied using several catalysts precipitated on pumice (67).

Esterification of Other Compounds

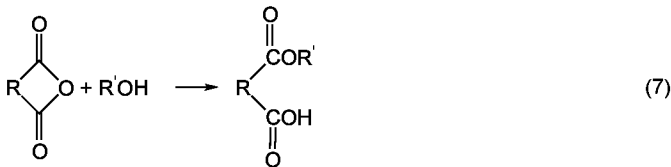
Acid Anhydrides. Acid anhydrides react with alcohols to form esters (in high yields in many cases) with a carboxylic acid formed as by-product:



However, this method is applied only when esterification cannot be effected by the usual acid–alcohol reaction because of the higher cost of the anhydrides. The production of cellulose acetate (see FIBERS, CELLULOSE ESTERS), phenyl acetate (used in acetaminophen production), and aspirin (acetylsalicylic acid) (see SALICYLIC ACID) are examples of the large-scale use of acetic anhydride. The speed of acylation is greatly increased by the use of catalysts (68) such as sulfuric acid, perchloric acid, trifluoroacetic acid, phosphorus pentoxide, zinc chloride, ferric chloride, sodium acetate, and tertiary amines, eg, 4-dimethylaminopyridine.

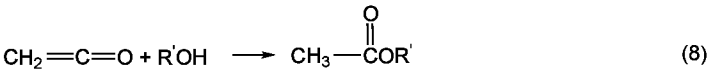
Formic anhydride is not stable. However, formate esters of alcohols and phenolics can be prepared using formic–acetic anhydride (69,70). Anhydrides can also be incorporated into polystyrene backbones which can then be treated with alcohols to afford the corresponding esters and carboxypolystyrene for recycle (71).

Dibasic acid anhydrides such as phthalic anhydride and maleic anhydride readily react with alcohols to form the monoalkyl ester:



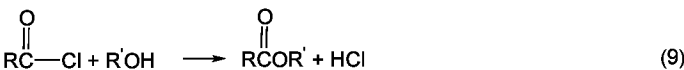
This reaction can be used for identification of individual alcohols because of the wide variations noted in the melting points of monoalkyl esters up to the dodecyl derivatives. The reaction can be used to separate alcohols of various classes. Monoesters are converted into the normal diesters by heating with an excess of alcohol and a catalyst; however, diesters are generally formed directly from the corresponding diacids.

Ketene, like acid anhydrides, reacts with alcohols to form (acetate) esters:

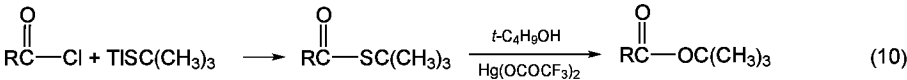


Ketene is an efficient acetylating agent with some alcohols, but in the absence of catalysts may be either nonreactive or sluggish with others, especially phenols and tertiary alcohols (72) (see KETENES AND RELATED SUBSTANCES).

Acid Chlorides. Acid chlorides react with alcohols to form esters:



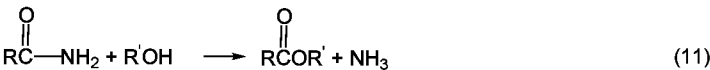
The acid chlorides are generally more reactive than the corresponding acid anhydrides. In fact, the alcoholysis of acid chlorides is probably the best laboratory method for preparing esters. Frequently, basic materials are added during the course of the reaction to neutralize by-product hydrochloric acid. When the basic material is aqueous caustic, the procedure is referred to as the Schotten-Baumann procedure (73). Esterification of tertiary alcohols by acid chlorides is described in Reference 74. Esters of tertiary alcohols can also be formed through an intermediate *t*-butyl thioate group (75):



Acid chlorides are used for the quantitative determination of hydroxyl groups and for acylation of sugars. Industrial applications include the formation of the alkyl or aryl carbonates from **phosgene** (see CARBONIC AND CHLOROFORMIC ESTERS) and phosphate esters such as triethyl, triphenyl, tricresyl, and tritolyl phosphates from phosphorus oxychloride.

The reaction of alcohols and acid chlorides in the presence of magnesium has been described (68). With primary and secondary alcohols the reaction is very smooth, and affords high and sometimes quantitative yields. Difficulty esterifiable hydroxy compounds such as tertiary alcohols and phenols can be esterified by this method. The reaction carried out in ether or benzene is usually very vigorous with evolution of hydrogen.

Amides. Alcoholysis of amides provides another method for synthesizing esters:



In order to produce high yields of ester in this manner it is necessary to remove the by-product ammonia (or amine) either by heating or combining with mineral acid, eg, H₂SO₄ or HCl. Recent work has shown that acidic ion-exchange resins can be used in place of mineral acids for converting sensitive unsubstituted amides (76). The structural relationships involved in esterification of amides are shown in Table 2 (77).

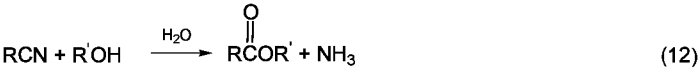
Table 2. Comparative Yields of Esters from Amides or Acids

Methyl ester	Yield of ester, %	
	From amide	From acid

formate	34	
acetate	70	56
monochloroacetate	64	65
dichloroacetate	57	70
trichloroacetate	53	73
phenylacetate	50	86
propionate	80	44
benzoate	15	37

Other methods of converting amides to esters have been described (78). Alkyl halides can be treated with amides to give esters (79). Also, esters can be synthesized from *N*-alkyl-*N*-nitrosoamides, which are derived from the corresponding amides (80).

Nitriles. Alcoholysis of nitriles offers a convenient way to produce esters without isolating the acid:



Acids are used to combine with the ammonia formed. A large excess of alcohol is used, but the amount of water is generally kept small. Catalysts such as hydrogen chloride, hydrogen bromide, and sulfuric acid have been employed (71).

One of the most important applications of this process is that of methyl methacrylate manufacture. In this process (81), acetone cyanohydrin is treated with sulfuric acid at 100°C, affording the corresponding methacrylamide sulfate which is esterified with methanol. After purification, methyl methacrylate (99.8% purity) is obtained in a yield of ca 85%.

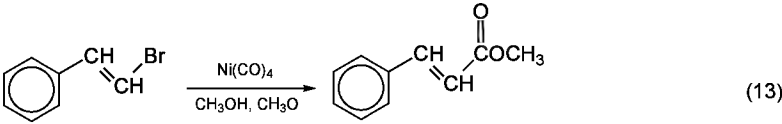
Unsaturated Hydrocarbons. Olefins from ethylene through octene have been converted into esters via acid-catalyzed nucleophilic addition. With ethylene and propylene, only a single ester is produced using acetic acid, ethyl acetate and isopropyl acetate, respectively. With the butylenes, two products are possible: *sec*-butyl esters result from 1- and 2-butylenes, whereas *tert*-butyl esters are obtained from isobutylene. The C5 olefins give rise to three *sec*-amyl esters and one *t*-amyl ester. As the carbon chain is lengthened, the reactivity of the olefin with organic acids increases.

In the case of ethylene, it is necessary to use high temperatures and pressures as well as active catalyst to effect esterification (82). Yields of 40–50% based on ethylene were obtained with boron trifluoride–hydrogen fluoride mixtures as catalysts at 150°C. 2-Butene under pressure at 115–120°C with an excess of glacial acetic acid containing 10% H₂SO₄ gave as much as a 60% yield of *sec*-butyl acetate (83).

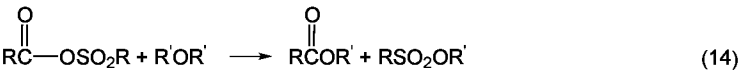
tert-Butyl acetate [540-88-5] was prepared by passing isobutylene and acetic acid (2:1 mol ratio) in the liquid phase over a silica catalyst impregnated with vanadium pentoxide and potassium sulfate at 1.7 MPa (250 psi). Conversion of isobutylene to ester increased with increasing temperature and ranged from 10% at 52°C to 24% at 93°C. Based on the acetic acid charged, yields of 31–43% of *t*-butyl acetate resulted at 93°C (84).

Most of the vinyl acetate produced in the United States is made by the vapor-phase ethylene process. In this process, a vapor-phase mixture of ethylene, acetic acid, and oxygen is passed at elevated temperature and pressures over a fixed-bed catalyst consisting of supported palladium (85). Less than 70% oxygen, acetic acid, and ethylene conversion is realized per pass. Therefore, these components have to be recovered and returned to the reaction zone. The vinyl acetate yield using this process is typically in the 91–95% range (86). Vinyl acetate can be manufactured also from acetylene, acetaldehyde, and the liquid-phase ethylene process (see VINYL POLYMERS).

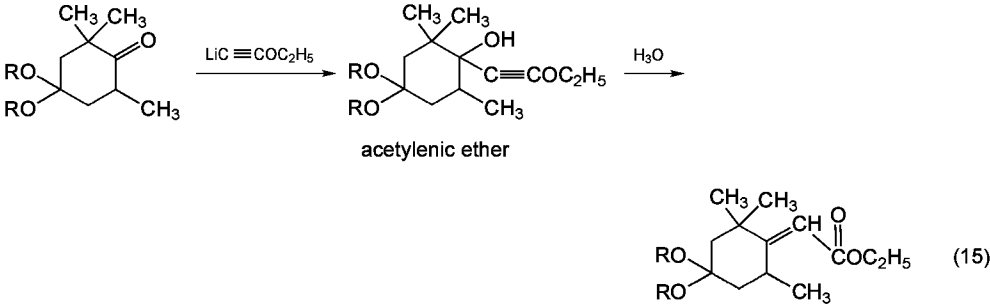
Esters can be obtained from halogenated olefins using a metal carbonyl catalyst (87), eg, *trans*-1-bromo-2-phenylethylene is treated with nickel carbonyl in the presence of methanol to afford the corresponding methyl cinnamate (see CINNAMIC ACID).



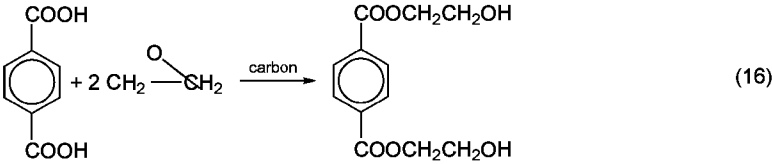
Ethers. In the presence of anhydrous agents such as ferric chloride (88), hydrogen bromide, and acid chlorides, ethers react to form esters (see ETHERS). Esters can also be prepared from ethers by an oxidative process (89). With mixed sulfonic–carboxylic anhydrides, ethers are converted to a mixture of the corresponding carboxylate and sulfonate esters (90):



Unsaturated esters can be prepared from the corresponding acetylenic ethers with yields in most cases of >50% (91) as in the following example:



β-hydroxyethyl esters can be prepared from carboxylic acids and ethylene oxide:

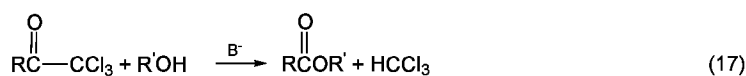


Bis-(β-hydroxyethyl) terephthalate and related compounds can be produced in this manner using finely divided carbon catalyst (92). The carbon functions not only as a catalyst but also helps to remove color from the reaction mixture upon removal of the carbon by hot filtration.

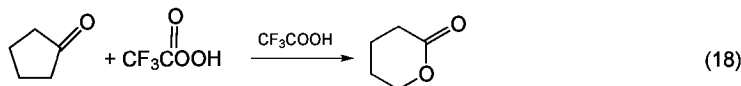
Aldehydes and Ketones. Esters are obtained readily by condensation of aldehydes in the presence of alcoholate catalysts such as aluminum

ethylate, $\text{Al}(\text{OC}_2\text{H}_5)_3$, by the Tishchenko reaction. The alcoholate catalysts may be prepared from commercial aluminum and *n*-butyl or isobutyl alcohol in the presence of 2–5% aluminum chloride (93).

Trihalomethyl ketones react with alcohols in the presence of alkaline catalysts even at room temperature (94):

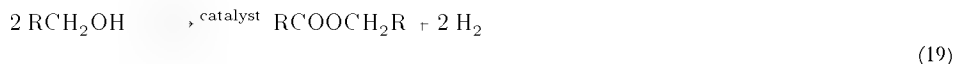


A variety of esters can be prepared from the corresponding ketones using peracids in a process usually referred to as the Baeyer-Villiger reaction (95); ie, cyclopentanone is converted to δ -valerolactone upon treatment of the ketone with peroxytrifluoroacetic acid:



This conversion can be carried out, in many cases, with >80% yield.

Alcohols. The direct synthesis of esters by dehydrogenation or oxidative hydrogenation of alcohols offers a simple method for the preparation of certain types of esters, such as ethyl acetate (96–98):



The reaction is catalyzed by copper with various promoters or activators, and is carried out in the vapor phase at 200–300°C.

Technical Preparation of Esters

Esterification is generally carried out by refluxing the reaction mixture until the carboxylic acid has reacted with the alcohol and the water has been split off. The water or the ester is removed from the equilibrium by distillation. The choice of the esterification process to obtain a maximum yield is dependent on many factors, ie, no single process has universal applicability. Although extensive preparative techniques have been reviewed elsewhere (7,68), the methods given in this section are representative of both laboratory and plant-scale techniques used in batch esterifications.

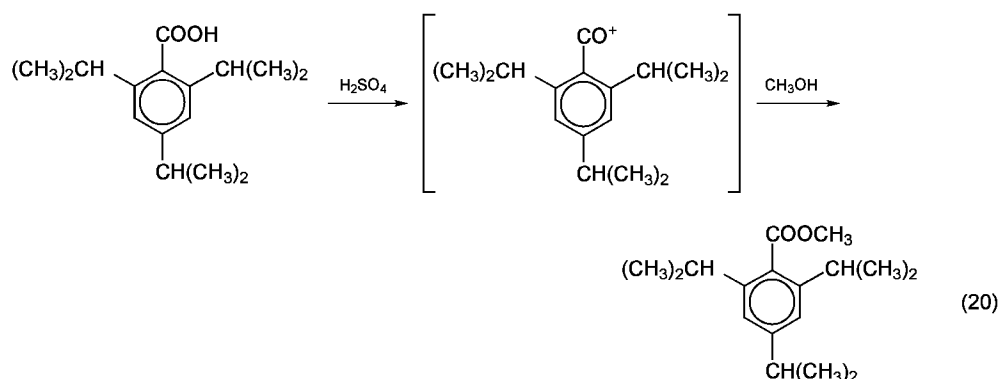
Methyl Esters. Methyl esters are obtained in good yield using methylene dichloride or ethylene dichloride as solvent (99). The latter is generally preferred, but the choice of the solvent depends to some extent on the boiling point of the desired ester. Also, the toxicity of these solvents should be considered prior to using them (see [CHLOROCARBONS AND CHLOROHYDROCARBONS](#)). The general procedure is as follows: for each mole of aliphatic carboxyl group, 96 g (3 mol) of methanol, 300 mL of ethylene dichloride, and 3 mL of concentrated H_2SO_4 are used. With aromatic acids, the amount of H_2SO_4 is increased to 15 mL/mol of carboxyl group. The mixture is refluxed for 6–15 h, although in some cases the time may be as short as 30 minutes. Progress of esterification is usually indicated by the development of cloudiness and separation of an upper layer containing water, methanol, and sulfuric acid. After the reaction is completed, the cooled mixture is washed successively with water, sodium bicarbonate solution, and again with water. The ethylene chloride layer is then distilled at atmospheric or reduced pressure, and the residual methyl ester is purified by distillation or crystallization.

The manufacture of high purity methyl acetate by a reactive distillation process has been accomplished; high conversion of one reactant can be achieved only with a large excess of the other reactant. Because the reaction is reversible, the rate of reaction in the liquid phase is increased by removing methyl acetate preferentially to the other components in the reaction mixture (100).

Medium Boiling Esters. Esterification of ethyl and propyl alcohols, ethylene glycol, and glycerol with various acids, eg, chloro- or bromoacetic, or pyruvic, by the use of a third component such as benzene, toluene, hexane, cyclohexane, or carbon tetrachloride to remove the water produced is quite common. Benzene has been used as a co-solvent in the preparation of methyl pyruvate from pyruvic acid (101). The preparation of ethyl lactate is described as an example of the general procedure (102). A mixture of 1 mol 80% lactic acid and 2.3 mol 95% ethyl alcohol is added to a volume of benzene equal to half that of the alcohol (ca 43 mL), and the resulting mixture is refluxed for several hours. When distilled, the overhead condensate separates into layers. The lower layer is extracted to recover the benzene and alcohol, and the water is discarded. The upper layer is returned to the column for reflux. After all the water is removed from the reaction mixture, the excess of alcohol and benzene is removed by distillation, and the ester is fractionated to isolate the pure ester.

High Boiling Esters. The following procedure can be used for making diethyl phthalate and other high boiling esters (103). Phthalic anhydride (1 equiv) and 2.5 equivalents of ethanol are refluxed for 2 h in the presence of 1% of concentrated H_2SO_4 . To produce the monoester, the excess of alcohol is distilled at <100°C. For the diester, a mixture of 67% benzene and 33% alcohol is introduced continuously below the surface of the reaction mixture and the resulting alcohol–water–benzene ternary is distilled and condensed. A yield of diester of >99% is obtained by passing 3.4–7 equivalents of alcohol through the mixture in 4.5–7 hours. In another continuous process for the production of diesters, the mixture of alcohol, acid, and the catalyst is introduced into the upper part of a distillation column and an excess of the alcohol is introduced into the bottom. The column is heated so that the ester, water, and excess alcohol are distilled off (104). Organotitanates, zirconates, or organotin compounds are effective catalysts for the esterification of carboxylic acids or anhydrides with higher boiling monohydroxy alcohols at temperatures that permit the continuous distillation of the water formed (105). Refluxing 1 mol phthalic anhydride with 3 mol 2-ethyl-2-hexanol with stirring using these agents, then removing the water by a trap separator gives the corresponding esters in ~99% yields (see [PHTHALIC ACID](#)). Phthalic anhydride has been esterified with >99% conversion with 10–30% excess alcohol in the presence of 0.8–1.5 mol% alkyl titanate containing 0.08–0.2% activating agent in a vertical, multistage reactor connected to a devolatilization column and filter (106).

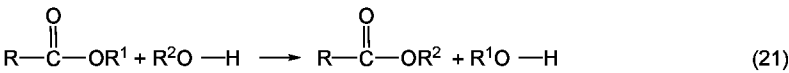
Difficulty Esterifiable Acids. The sterically hindered acids, such as 2,6-disubstituted benzoic acids, cannot usually be esterified by conventional means. Several esters of sterically hindered acids such as 2,4,6-triisopropylbenzoic acid [49623-71-4] have been prepared by dissolving 2 g of the acid in 14–20 mL of 100% H_2SO_4 (107). After standing a few minutes at room temperature, when presumably the acylium cation is formed (eq. 20), the solution is poured into an excess of cold absolute methanol. Most of the alcohol is removed under reduced pressure, about 50 mL of water is added, and the distillation is continued under reduced pressure to remove the remainder of the methanol. The organic matter is extracted with ether and treated with sodium carbonate solution. The ester is then distilled. Yields of esters made in this manner are 57–81%.



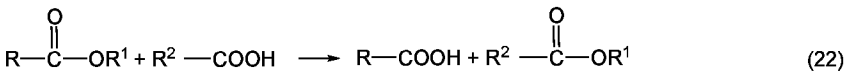
Ester Interchange

Ester interchange (transesterification) is a reaction between an ester and another compound, characterized by an exchange of alkoxy groups or of acyl groups, and resulting in the formation of a different ester. The process of transesterification is accelerated in the presence of a small amount of an acid or a base.

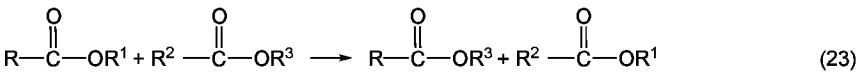
Three types of transesterification are known: (1) exchange of alcohol groups, commonly known as alcoholysis. In this process the compound with which an ester reacts is an alcohol:



(2) exchange of acid groups, acidolysis. In this process the compound with which an ester reacts is an acid:



(3) ester–ester interchange. In this process an exchange takes place between two esters:



These reactions are reversible and ordinarily do not involve large energy changes.

ESTER–ALCOHOL INTERCHANGE

Reaction Conditions. Alcoholysis commonly takes place in one liquid phase, sometimes with one of the reactants being only partially soluble and going into solution gradually as the reaction proceeds. Unless an excess of one of the reactants is used, or unless one of the products is withdrawn from the reaction phase by vaporization or precipitation, the reaction does not proceed to completion but comes to a standstill with substantial proportions of both alcohols and both esters in equilibrium. The concentrations present at equilibrium depend on the characteristics of the alcohols and esters involved, but in most practical uses of the reaction, one or both of the devices mentioned are used to force the reaction toward completion.

Temperatures. With alkaline catalysts, the reaction often takes place at RT or even lower temperatures. With acid catalysts, temperatures near 100°C are commonly used. With no catalyst, temperatures ~250°C may be required for a practical reaction rate.

Catalysts. Of the alkaline catalysts, alkali metal alkoxides are the most effective; ordinarily, the sodium or potassium alkoxide of the alcohol entering the reaction is preferred. Various other catalysts of milder alkalinity are preferred in special cases. For example, the use of sodium methyl carbonate as catalyst in the methanolysis of poly(vinyl acetate) is said to yield a poly(vinyl alcohol) having improved color. Aluminum alkoxide has been proposed as a catalyst for the alcoholysis of certain unsaturated esters; other sensitive esters have been made with a Grignard reagent as catalyst. Zinc is reported to be an efficient catalyst in the alcoholysis of ethyl esters of α-halogenated aliphatic acids by allyl and methallyl alcohols; conventional catalysts favor undesirable side reactions. Neutral organic titanates have received much attention (108). Divalent metal salts such as zinc or manganese acetate and organotinns such as dibutyltin oxide have been employed.

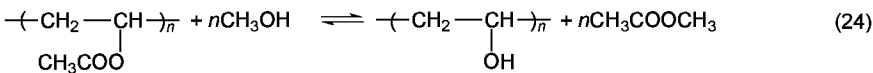
Among the acid catalysts, sulfuric acid, sulfonic acids, and hydrochloric acid are most used. With polyhydric alcohols, sulfuric acid is preferred to hydrochloric acid because of the tendency of hydrochloric acid to form chlorohydrins.

Equilibrium. In general, primary alcohols are more reactive than secondary alcohols (that is, they tend to displace them), and secondary alcohols tend to displace tertiary alcohols, but in addition, there are considerable differences among different members of the same class. Various alcohols have been compared in this way (4,109).

Applications. Transesterifications via alcoholysis play a significant role in industry as well as in laboratory and in analytical chemistry. The reaction can be used to reduce the boiling point of esters by exchanging a long-chain alcohol group with a short one, eg, methanol, in the analysis of fats, oils, and waxes. For more details see References 7 and 68. A few examples are given below.

***n*-Butyl Oleate.** Olive oil, 3 kg, consisting mainly of the glyceryl esters of oleic acid, is refluxed for 20 h with 7 L of *n*-butyl alcohol containing 150 g of concentrated H₂SO₄. The product contains a small proportion of saturated esters (110).

Poly(vinyl alcohol). Poly(vinyl alcohol) (see VINYL POLYMERS) is more easily prepared, in a form that can be filtered and washed in a practical way, by alcoholysis of poly(vinyl acetate), than by its saponification in an aqueous system:



The use of a catalytic quantity of alkali equivalent to only a small fraction of the acetate has the advantage that contamination of the poly(vinyl alcohol) with salts, which are difficult to remove, is minimized. A variant of the process is the use of a mixture of alcohol with the acetate ester produced by the alcoholysis as the alcoholyzing agent. This provides a means of controlling the completeness of removal of the acetate groups from the poly(vinyl acetate) (111).

Acrylic Esters. A procedure has been described for preparation of higher esters from methyl acrylate that illustrates the use of an acid catalyst together with the removal of one of the products by azeotropic distillation (112). Another procedure for the preparation of butyl acrylate, secondary alkyl acrylates, and hydroxyalkyl acrylates using *p*-toluenesulfonic acid as a catalyst has been described (113). Aluminum isopropoxide catalyzes the reaction of amino alcohols with methyl acrylate and methyl methacrylate. A review of the synthesis of acrylic esters by transesterification is given in Reference 114 (see ACRYLIC ACID AND DERIVATIVES).

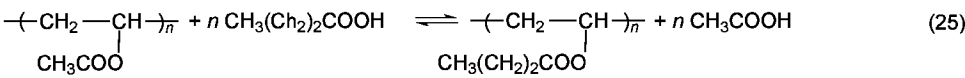
ESTER–ACID INTERCHANGE

Acidolysis requires the sue of an elevated temperature, the use of an acid catalyst (7), or both. Like alcoholysis, the reaction is reversible and requires the use of an excess of the replacing acid or removal of one of the products from the reaction if a high degree of replacement of the acid radical of an ester by another acid is to be obtained. This can be accomplished by distilling one of the products from the reaction mixture during the acidolysis.

In a series of organic acids of similar type, not much tendency exists for one acid to be more reactive than another. For example, in the replacement of stearic acid in methyl stearate by acetic acid, the equilibrium constant is 1.0. However, acidolysis in formic acid is usually much faster than in acetic acid, due to higher acidity and better ionizing properties of the former (115). Branched-chain acids, and some aromatic acids, especially sterically hindered acids such as ortho-substituted benzoic acids, would be expected to be less active in replacing other acids. Mixtures of esters are obtained when acidolysis is carried out without forcing the replacement to completion by removing one of the products. The acidolysis equilibrium and mechanism are discussed in detail in Reference 115.

An industrial example of acidolysis is the reaction of poly(vinyl acetate) with butyric acid to form poly(vinyl butyrate). Often a butyric acid–methanol

mixture is used and methyl acetate is obtained as a coproduct.



ESTER–ESTER INTERCHANGE

The reaction between two esters to produce two other esters was described by Friedel and Crafts in 1865, but has not been used as much as alcoholysis. The same general principles apply with regard to reversibility of the reaction and the means of driving the reaction to completion (7). In general, the same catalysts are effective as in alcoholysis. Usually the reaction is slower than alcoholysis of the same esters. Without a catalyst, a reaction time of several h at >250° C is required to bring two typical esters to equilibrium. Catalysts are almost essential to bring reaction rates into a practical range so that the use of destructive temperatures can be avoided. Tin compounds, especially stannous hydroxide, have been mentioned frequently as catalysts and do not produce much decomposition or discoloration of the esters (116). More effective at lower temperatures are the acid catalysts, such as sulfuric acid and sulfonic acids, and especially the alkaline catalysts such as sodium alkoxides. With an alkaline catalyst, ester–ester interchange can be carried out at temperatures as low as 0°C.

BIBLIOGRAPHY

"Esterification" in *ECT* 1st ed., Vol. 5, pp. 776–817, by C. E. Leyes, Celanese Corp. of America; "Ester Interchange" in *ECT* 1st ed., Vol. 5, pp. 817–823, by E. W. Eckey, E. W. Eckey Research Laboratory; "Esterification" in *ECT* 2nd ed., Vol. 8, pp. 313–356, by C. E. Leyes, Newark College of Engineering; "Ester Interchange" in *ECT* 2nd ed., Vol. 8, pp. 356–365 by E. W. Eckey, E. W. Eckey Research Laboratory, and E. F. Izard, E. I. du Pont de Nemours & Co., Inc., "Esterification" in *ECT* 3rd ed., Vol. 9, pp. 291–310, by E. G. Zey, Celanese Corp.

1. H. A. Goldsmith, *Chem. Rev.* **33**, 257 (1943).
2. D. B. Keyes, *Ind. Eng. Chem.* **24**, 1096 (1932).
3. E. E. Reid in P. Grotggins, *Unit Processes in Organic Synthesis*, 5th ed., McGraw-Hill Book Co., Inc., New York, 1958.
4. S. Patai, *The Chemistry of Carboxylic Acids and Esters*, Wiley-Interscience, New York, 1969.
5. H. H. Szmant, *Organic Building Blocks of the Chemical Industry*, Wiley-Interscience, New York, 1989.
6. M. L. Bender, *Chem. Rev.* **60**, 53 (1960).
7. K. S. Markley in K. S. Markley, ed., *Fatty Acids*, part 2, Wiley-Interscience, New York, 1961, p. 757.
8. J. March, *Advanced Organic Chemistry*, 3rd ed., John Wiley & Sons, Inc., New York, 1985.
9. R. C. Larock, *Comprehensive Organic Transformations*, VCH Publishers, Inc., New York, 1989.
10. *Chem. Eng. News* **69**, 28 (June 24, 1991).
11. B. Holmberg, *Chem. Ber.* **45**, 2997 (1912).
12. M. Polanyi and A. L. Szabo, *Trans. Faraday Soc.* **30**, 508 (1934).
13. E. K. Euranto in Ref. 4, p. 505.
14. R. C. Sharma and M. M. Sharma, *Bull. Chem. Soc. (Japan)* **43**, 642 (1970).
15. R. C. Sharma and M. M. Sharma, *J. Appl. Chem.* **19**, 162 (1969).
16. R. J. Hartman, L. B. Storms, and A. G. Gassmann, *J. Am. Chem. Soc.* **61**, 2167 (1939) and references therein.
17. A. G. Gassmann and R. J. Hartman, *J. Am. Chem. Soc.* **63**, 2393 (1941) and references therein.
18. M. Berthelot and L. Pean de Saint-Gilles, *Ann. Chim. Phys.* **68**, 225 (1863).
19. E. S. Gould, *Mechanisms and Structure in Organic Chemistry*, Holt, Rinehard, and Winston, New York, 1959, p. 314.
20. J. Hine, *Physical Organic Chemistry*, McGraw-Hill Book Co., Inc., New York, 1962, p. 275.
21. C. A. Lane, *J. Am. Chem. Soc.* **86**, 2521 (1964).
22. Y. K. Syrkin and I. I. Moiseev, *Usp. Khim.* **27**, 717 (1958).
23. V. A. Palm and co-workers, *Zh. Fiz. Khim.* **36**, 2499 (1962).
24. C. Murphy and co-workers, *Chem. Petro-Chem. J.* **10**(5), 13 (1979).
25. J. L. Guy, *Chem. Eng.* **89** (13), 97 (1982).
26. M. B. Bogacki and J. Szymanowski, *Pr. Nauk. Inst. Chem. Nieorg. Metal. Pierwiotkow Rządskich Politech. Wrocław.* **55**, 99 (1986); *Chem. Abstr.* **106**, 20447m (1986).
27. T. Yamada and Y. Imamura, *Polym.-Plast. Technol. Eng.* **28**(7–8), 811 (1989).
28. V. Parausanu and co-workers, *Rev. Chim. (Bucharest)* **31**(1), 30 (1980).
29. N. Schlechter, D. F. Othmer, and S. Marshak, *Ind. Eng. Chem.* **37**, 900 (1945).
30. C. E. Leyes and D. F. Othmer, *Ind. Eng. Chem.* **37**, 968 (1945).
31. S. Berman, A. A. Melnychuk, and D. F. Othmer, *Ind. Eng. Chem.* **40**, 1312 (1948).
32. M. Berthelot and P. Saint-Gilles, *Ann. Chim. Phys.* **65**, 385 (1862).
33. W. Swietoslawski, *J. Phys. Chem.* **37**, 701 (1933).
34. P. E. Coria, *Rev. Fac. Cienc. Quim. Univ. Nac. La Plata* **10**, 67 (1935).
35. C. A. Durruty, *An. Assoc. Quim. Argent.* **19**, 227 (1931).
36. H. M. Trimble and E. L. Richardson, *J. Am. Chem. Soc.* **62**, 3721 (1932).
37. W. P. Jencks and M. Gilchrist, *J. Am. Chem. Soc.* **86**, 4651 (1964).
38. **Jpn. Pat. 62,096,453** (May, 1987), M. Hirata, K. Ishikawa, and S. Naruto (to Japan Organo Co., Ltd.).
39. L. H. Horsley in R. F. Gould, ed., *Azotropic Data—III*, American Chemical Society, Washington, D.C., 1973.
40. J. C. Konen, E. T. Clocker, and R. P. Cox, *Oil Soap* **22**, 57 (1945).
41. E. Thielpape and A. Fulde, *Chem. Ber.* **B66**, 1454 (1933).
42. P. L. Gordon and R. Aronowitz, *Ind. Eng. Chem.* **37**, 780 (1945).
43. I. T. Harrison and S. Harrison, *Compendium of Organic Synthetic Methods*, Vol. 1, Wiley-Interscience, New York, 1971.
44. I. T. Harrison and S. Harrison, *Compendium of Organic Synthetic Methods*, Vol. 2, Wiley-Interscience, New York, 1974.
45. L. S. Hegedus and L. Wade, *Compendium of Organic Synthetic Methods*, Vol. 3, Wiley-Interscience, New York, 1977.
46. **U.S. Pat. 4,833,267** (1989), S. Nakashima and co-workers (to Nippon Shokubai Kagaku Kogyo Co.).
47. S. S. Bhagade and G. D. Nageshwar, *Chem. Petro-Chem. J.*, **9**(7), 3 (1978).
48. D. J. Pietrzyk, *Chromatogr. Sci.* **47**, 585 (1990).
49. Z. Gomzi and S. Zrnecvic, *Croat. Chem. Acta.* **53**(1), 25 (1980).
50. *Chem. Abstr.* **110**, 137443f (1989).
51. WO 9,008,121 (1990), M. Wilmott and co-workers (to Davy McKee (London) Ltd.).
52. C. L. Levesque and A. M. Craig, *Ind. Eng. Chem.* **40**, 96 (1948).
53. M. J. Astel, B. Schaeffer, and C. O. Obenland, *J. Am. Chem. Soc.* **77**, 3643 (1955).

54. R. Neumann and Y. Sasson, *Ind. Eng. Chem. Process Des. Dev.* **23**, 654 (1984).

55. C. E. Leyes and D. F. Othmer, *Trans. Am. Inst. Chem. Eng.* **41**, 157 (1945).

56. **U.S. Pat. 2,426,968** (Sept. 2, 1947), H. W. Grubb, L. M. O'Hara, and K. Atwood (to Seagram and Sons).

57. **U.S. Pat. 4,374,070** (1983), T. Larkin, S. Polichnowski, and G. Tustin (to Eastman Kodak Co.).

58. **U.S. Pat. 4,435,595** (1984), V. Agreda and L. Partin (to Eastman Kodak Co.).

59. V. Agreda, L. Partin, and W. Heise, *Chem. Eng. Prog.* **40** (1990) and references therein.

60. **Brit. Pat. 2,033,385** (1980), C.-G. Wan (to Halcon Research and Development Corp.).

61. **Eur. Pat. 87,870** (1983), J. B. Cooper (to British Petroleum Chem. Ltd.).

62. V. I. Goldanskii, *J. Phys. Chem. USSR* **21**, 431 (1947); H. C. Tidwell and E. E. Reid, *J. Am. Chem. Soc.* **53**, 4353 (1931).

63. A. Mailhe and F. deGodon, *Bull. Soc. Chim. Fr.* **29**, 101 (1921); G. Edgar and W. H. Schuyler, *J. Am. Chem. Soc.* **46**, 64 (1924).

64. H. F. Hoerig, D. Hanson, and O. L. Kowalke, *Ind. Eng. Chem.* **35**, 575 (1943).

65. A. M. Chashchin and N. M. Levedeva, *Gidroliz Lesokhim. Promst.* **15**, 6 (1962).

66. S. Sharipova, A. Arifdzhanov, and A. Sultanov, *Katal. Pererab. Uglevodorodnogo Syr'ya*, 108 (1967).

67. J. F. Spangenberg, *Ind. Quim. Buenos Aires* **7**, 393 (1945).

68. C. A. Buehler and D. E. Pearson, *Survey of Organic Synthesis*, Wiley-Interscience, New York, 1970, p. 801.

69. A. van Es and W. Stevens, *Recl. Trav. Chim. Pays-Bas* **84**, 704 (1965).

70. S. Sofuku and co-workers, *Bull. Chem. Soc. Jpn.* **40**, 2942 (1967).

71. M. B. Shambhu and G. A. Digenis, *Tetrahedron Lett.*, 1627 (1973).

72. C. D. Hurd and A. S. Roe, *J. Am. Chem. Soc.* **61**, 3355 (1939).

73. N. O. V. Sonntag, *Chem. Rev.* **52**, 272 (1953).

74. W. H. Puterbaugh, *J. Org. Chem.* **27**, 4010 (1962).

75. S. Masamune and co-workers, *J. Am. Chem. Soc.* **97**, 3515 (1975).

76. W. J. Greenlee and E. D. Thorsett, *J. Org. Chem.* **46**, 5351 (1981).

77. S. G. Toole and F. J. Sowa, *J. Am. Chem. Soc.* **59**, 1971 (1937).

78. Ref. 43, p. 289.

79. J. S. Matthews and J. P. Cookson, *J. Org. Chem.* **34**, 3204 (1969).

80. E. H. White, *J. Am. Chem. Soc.* **77**, 6011 (1955).

81. Ref. 5, pp. 86 and 312.

82. J. A. John in S. A. Miller, ed., *Ethylene and Its Industrial Derivatives*, Ernest Benn Ltd., London, 1969, p. 765.

83. B. T. Brooks, *Ind. Eng. Chem.* **27**, 278 (1935).

84. **U.S. Pat. 3,014,066** (Dec. 19, 1961), E. R. Kerr and M. C. Throckmorton (to Texaco).

85. W. Schwerdtel, *Hydrocarbon Process.* **47**, 187 (1968).

86. **Brit. Pat. 981,987** (Feb. 3, 1955), (to Farbenfabriken Bayer AG).

87. E. J. Corey and L. S. Hegedus, *J. Am. Chem. Soc.* **91**, 1233 (1969).

88. B. Ganem and V. R. Small, *J. Org. Chem.* **39**, 3728 (1974).

89. E. C. Juenge and D. A. Beal, *Tetrahedron Lett.*, 5819 (1968).

90. M. H. Karger and Y. Mazur, *J. Org. Chem.* **36**, 532 (1971).

91. J. N. Marx and J. Sondheimer, *Tetrahedron (suppl. 8, part 1)*, 1 (1966).

92. **U.S. Pat. 3,652,647** (Mar. 28, 1972), E. G. Zey (to Celanese).

93. M. Y. Kagen and I. A. Sobolev, *J. Chem. Ind. (Moscow)* (2), 35 (1933).

94. J. Houben and W. Fisher, *Chem. Ber.* **B64**, 240 and 2636 (1931).

95. H. O. House, *Modern Synthetic Reactions*, 2nd ed., W. A. Benjamin, Inc., Menlo Park, Calif., 1972, pp. 321–326.

96. Ref. 82, p. 764.

97. H. Gilman, ed., *Organic Synthesis*, Vol. 1, John Wiley & Sons, Inc., New York, 1932, pp. 138–140.

98. S. O. Nwauka and P. M. Keehn, *Tetrahedron Lett.* **23**, 35 (1982).

99. R. O. Clinton and S. C. Laskowski, *J. Am. Chem. Soc.* **70**, 3135 (1948).

100. **U.S. Pat. 4,939,294** (July 3, 1990), V. H. Agreda and R. D. Lilly (to Eastman Kodak Co.).

101. A. Weissberger and C. J. Kibler, *Organic Synthesis Coll. Vol. 3*, John Wiley & Sons, Inc., New York, 1955, p. 610.

102. **U.S. Pat. 1,421,604** (July 4, 1922), J. A. Steffens (to U.S. Industrial Alcohol Co.).

103. **U.S. Pat. 2,076,111** (Apr. 6, 1927), W. J. Bannister (to Commercial Solvents Corp.).

104. **Ger. Pat. DE 878,348** (1942), H. Dierichs and co-workers (to BASF).

105. **Brit. Pat. 852,110** (Oct. 26, 1960), (to B. F. Goodrich Co.).

106. **Rom. Pat. 66,492B** (Oct. 20, 1979), I. Drimus and E. D. Morath (Combinatul Petrochemic "Solventul").

107. M. S. Newman, *J. Am. Chem. Soc.* **63**, 2431 (1941).

108. *TYZOR Organic Titanates, Du Pont Bulletin D5258*, E. I. du Pont de Nemours & Co., Inc., Wilmington, Del., 1972, 5M. Rev. 377.

109. G. B. Hatch and J. E. Adkins, *J. Am. Chem. Soc.* **59**, 1694 (1937).

110. E. E. Reid and co-workers in A. H. Blatt, ed., *Organic Synthesis*, Coll. Vol. 2, John Wiley & Sons, Inc., New York, 1943, p. 469.

111. **U.S. Pat. 2,266,996** (Dec. 23, 1941), N. D. Scott and J. E. Bristol (to E.I. du Pont de Nemours & Co., Inc.).

112. C. E. Rehberg and C. H. Fisher, *J. Am. Chem. Soc.* **66**, 1203 (1944).

113. C. E. Rehberg, *Organic Synthesis*, Coll. Vol. 3, John Wiley & Sons, Inc., New York, 1955, p. 146.

114. J. K. Haken, *Synthesis of Acrylic Esters by Transesterification*, Noyes Development Corp., Park Ridge, N.J., 1967.

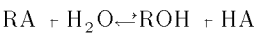
115. J. Koskikallio in S. Patai, ed., *The Chemistry of Carboxylic Acids and Esters*, Wiley-Interscience, New York, 1969, pp. 103–136.

116. A. E. Bailey, *Industrial Oil and Fat Products*, Interscience Publishers, New York, 1945, p. 676.

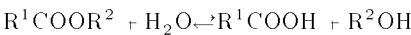
Mohammad Aslam
G. Paull Torrence
Edward G. Zey
Hoechst Celanese Corporation

ESTERS, ORGANIC

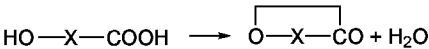
Esters are compounds that, on hydrolysis, yield alcohols or phenols and acids according to the equation:



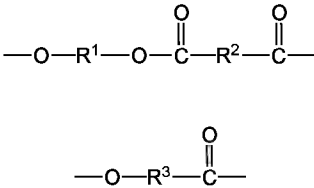
where R is a hydrocarbon fragment and A is the anion portion of an organic acid. For carboxylic acid esters, the reaction can be represented as:



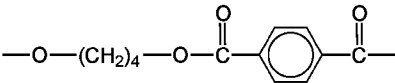
where R¹ and R² are the same or different hydrocarbon radicals. The reverse reaction constitutes the usual method for preparing esters (see ESTERIFICATION). When R¹ and R² are bonded together, the resultant cyclic ester is called a lactone. Lactones can be produced from molecules containing both carboxyl and hydroxyl groups.



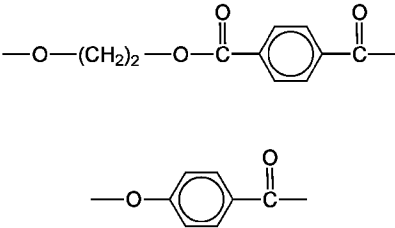
Polyesters are polymers with basic structural units:



where R¹, R², and R³ are alkyl, aromatic, or alkyl–aromatic radicals and they may be the same or different. For example, poly(butylene terephthalate) [26062-94-2] has the repeating unit



whereas poly(ethylene terephthalate–*p*-oxybenzoate) [25822-54-2] copolymer consists of the following two repeating units



Orthoesters, RC(OR¹)₃ (1), thioesters, RCSOR¹ (2,3) (see SULFUR COMPOUNDS; THIOLS), and carbamates, H₂NCOOR, are not covered in this review.

Nomenclature

The names of esters consist of two words that reflect their formation from an alcohol and a carboxylic acid. According to the IUPAC rule, the alkyl or aryl group of the alcohol is cited first followed by the carboxylate group of the acid with the ending -ate replacing the -ic of the acid (4,5). For example, CH₃CH₂COOCH₃, the methyl ester of propanoic acid, is called methyl propanoate [554-12-1] (or methyl propionate, if the trivial name, propionic acid, is used for the carboxylic acid). The monoesters of dibasic acids are named by inserting the word hydrogen between names of the alcohol and the carboxylate. The monomethyl ester of succinic acid, CH₃OCOCH₂CH₂COOH, is called methyl hydrogen succinate or more systematically methyl hydrogen butanedioate [3878-55-5].

Based on the IUPAC rule, esters of polyhydric alcohols with monobasic acids are named analogously to simple esters: 1,2-ethanediyl diacetate for ethylene glycol diacetate [111-55-7], 2-hydroxyethyl acetate for ethylene glycol monoacetate [542-59-6], 1,2,3-propanetriyl triacetate for glycerol triacetate [102-76-1]. Cyclic esters are called lactones, and are named by changing the -ic acid of the hydroxy-acid to -olactone. A Greek letter (α,bgr,γ,δ, etc) is used to designate the carbon atom that bears the hydroxyl group of the parent acid. Lactones are best named, however, as heterocyclic compounds. For

example, $\overbrace{\text{CH}_2\text{CH}_2\text{CH}_2\text{COO}}^{\text{cyclic}}$, γ-butyrolactone, is dihydro-2(3*H*)-furanone [96-48-0].

When the ester function is named as a substituent, it is indicated by alkoxycarbonyl or acyloxy depending on the connection to the —C=O group:

$\text{CH}_3\overset{\overset{\text{O}}{\parallel}}{\text{C}}\text{---R}$ is a methoxycarbonyl derivative whereas $\text{CH}_3\overset{\overset{\text{O}}{\parallel}}{\text{C}}\text{O---R}$ is the acetoxy derivative. In naming esters containing one or more substituents, it is necessary to indicate specifically in which portion of the molecule the substituents occur, eg, ClCH₂COOCH₂CH₃ is ethyl chloroacetate [105-39-5] and CH₃COOCH₂CH₂Cl is 2-chloroethyl acetate [542-58-5].

Orthoesters are trivially named as derivatives of ortho acids such as triethyl orthoformate [122-51-0], HC(OC₂H₅)₃, or named systematically as ethers, 1,1,1-triethoxymethane.

Physical Properties

The physical properties of organic esters vary according to the molecular weight of each component (6–10). Lower molecular weight esters are colorless, mobile, and highly volatile liquids that usually have pleasant odors. As the molecular weight increases, volatility decreases and the consistency becomes waxy, then solid, and eventually even brittle, often with formation of lustrous crystals. The melting point of an ester is generally lower than that of the corresponding carboxylic acid. However, the boiling point depends on the chain length of the alcohol component and eventually exceeds that of the acid. Lower molecular weight esters are relatively stable when dry and can be distilled without decomposition. Organic esters are generally insoluble in water, but soluble in various organic liquids. Lower esters are themselves good solvents for many organic compounds. The physical properties of commercially important aliphatic and aromatic organic esters are listed in Table 1.

Table 1. Physical Properties of Some Common Esters

Ester	CAS Registry Number	Mol wt	n_D^{20}	d_{20}^{20}	Bp, °C ^a	Freezing point, °C	Flash point, °C ^b
methyl formate	[107-31-3]	60.05	1.344	0.0975	32	−99.8	−19
ethyl formate	[109-94-4]	74.08	1.3598	0.9236	54.3	−80	−20
butyl formate	[592-84-7]	102.13	1.3889	0.8885 ^c	106	91.9	18
methyl acetate	[79-20-9]	74.08	1.3594	0.933	57	98.1	10
ethyl acetate	[141-78-6]	88.1	1.3723	0.0902	77.1	83.6	4
vinyl acetate	[108-05-4]	86.1	1.3959	0.932	72.2	93.2	8
propyl acetate	[109-60-4]	102.13	1.3844	0.887	101.6	92.5	13
isopropyl acetate	[108-21-4]	102.13	1.3773	0.872	90	73.4	2
butyl acetate	[123-86-4]	116.16	1.3951	0.882	126	73.5	22
isobutyl acetate	[110-19-0]	116.16	1.3902	0.871	117.2	−98.6	18
sec-butyl acetate	[105-46-4]	116.16	1.3877	0.8758 ^d	112		31.1 ^e
<i>t</i> -butyl acetate	[540-88-5]	116.16	1.3855	0.8665 ^c	97		
pentyl acetate	[628-63-7]	130.18	1.4023	0.876	149.3	−70.8	25
isoamyl acetate	[123-92-2]	130.18	1.4000	0.872	142	78	25
sec-hexyl acetate	[108-84-9]	144.22	1.4014 ^f	0.8651 ^g	157	0	
2-ethylhexyl acetate	[103-09-3]	172.26	1.4204	0.873	199.3	93	71
ethylene glycol diacetate	[111-55-7]	146.14	1.415	1.128	191	31	88
2-methoxyethyl acetate	[110-49-6]	118.13	1.4019	1.0067	145	−65.1	44
2-ethoxyethyl acetate	[111-15-9]	132.16	1.4058	0.975	156.4	61.7	47
2-butoxyethyl acetate	[112-07-2]	160.12	1.42	0.943	187.8	32	81
2-(2-ethoxyethoxy)ethyl acetate	[111-90-0]	176.21	1.423	1.011	217.4	−25	107
2-(2-butoxyethoxy)ethyl acetate	[112-34-5]	204.27	1.4265	0.981	247	−32.2	110
benzyl acetate	[140-11-4]	150.18	1.5232	1.055	215.5	51.5	90
glyceryl triacetate	[102-76-1]	218.23	1.4296	1.161	258	78	138
ethyl 3-ethoxypropionate	[763-69-9]	146.19		0.95	165–172	−50	58
glyceryl tripropionate	[139-45-7]	260.3	1.4318	1.100 ^h	176	58	167 ^e
methyl acrylate	[96-33-3]	86.09	1.4040	0.953	80.5	< 75	3
ethyl acrylate	[140-88-5]	100.11	1.4068	0.923	99.8	< 72	10
butyl acrylate	[141-32-2]	128.17	1.4185	0.898	69	64.6	29
2-ethylhexyl acrylate	[103-11-7]	184.28		0.887	130 ⁱ	−90	82 ^e
methyl methacrylate	[80-62-6]	100.12	1.4119	0.944	100	−48	10 ^e
methyl butyrate	[623-42-7]	102.13	1.3878	0.898	102.3	84.8	14
ethyl butyrate	[105-54-4]	116.16	1.4000	0.878	121.6	100.8	24
butyl butyrate	[109-21-7]	144.22	1.4075	0.871	166.6	91.5	53
methyl isobutyrate	[547-63-7]	102.13	1.3840	0.891	92.6	84.7	
ethyl isobutyrate	[97-62-1]	116.16	1.3870	0.869	110	88	<21
isobutyl isobutyrate	[97-85-8]	144.22	1.3999	0.875	148.7	80.7	38
methyl stearate	[112-61-8]	298.5	1.457	0.836	215	40	153
ethyl stearate	[111-61-5]	312.52	1.429	1.057	213–215	33.7	
butyl stearate	[123-95-5]	340.58		0.855	343	27.5	160
dodecyl stearate	[5303-25-3]	440.8	1.433			28	
hexadecyl stearate	[1190-63-2]	496.91	1.441			57	
dimethyl maleate	[624-48-6]	144.13	1.4409	1.152	204		91
dimethyl oxalate	[95-92-1]	111.09	1.4096	1.148	185	41	76
dimethyl adipate	[627-93-0]	174.2	1.4283	1.0600	115	10.3	
diethyl adipate	[141-28-6]	202.25	1.4372	1.008	245	19.8	
di(2-ethylhexyl) adipate	[103-23-1]	370.58	1.4472	0.927	214	60	206
methyl benzoate	[93-58-3]	136.15	1.517	1.094	199.5	12.5	83
ethyl benzoate	[93-89-0]	150.18	1.505	1.051	212.9	34.2	88
methyl salicylate	[119-36-8]	152.15	1.536	1.184	223.3	8.6	96
ethyl salicylate	[118-61-6]	166.18	1.522	1.137	231.5	1.3	107
dimethyl phthalate	[131-11-3]	194.19	1.515	1.190	282	2	146
diethyl phthalate	[84-66-2]	222.24	1.499	1.118	295	33	161
dibutyl phthalate	[84-74-2]	278.35	1.4911	1.0465	340	35	157
di(2-ethylhexyl) phthalate	[117-81-7]	390.56	1.486	0.9861	231 ^j	50	218.3
dimethyl isophthalate	[1459-93-4]	194.19	1.5168	1.194 ^c	124	67	138
dimethyl terephthalate	[120-61-6]	194.19			288	140	153
methyl anthranilate	[134-20-3]	151.17	1.584	1.168	132	24	>100
benzyl cinnamate	[103-41-3]	238.29		1.109 ^g	244 ^j	39	110
dimethyl carbonate	[616-38-6]	90.08	1.3682	1.0694 ^c	90	3	19 ^e
diethyl carbonate	[105-58-8]	118.13	1.3854	0.9752 ^c	127	43	25

^a At 101.3 kPa (760 mm Hg unless otherwise stated).
^b Closed cup determination unless otherwise stated.

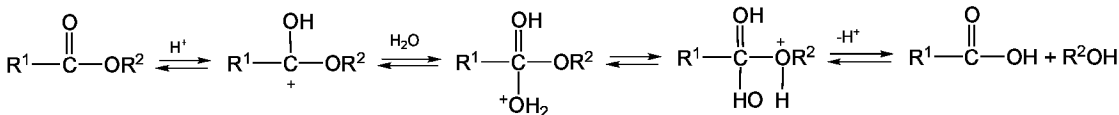
^c d_{20}^{20}

^d d_{4}^{25}
^e Open cup determination.
^f n_{D}^{25}
^g d_{15}^{15}
^h d_{18}^{20}
¹ At 6.7 kPa (50 mm Hg).
¹ At 0.67 kPa (5 mm Hg).

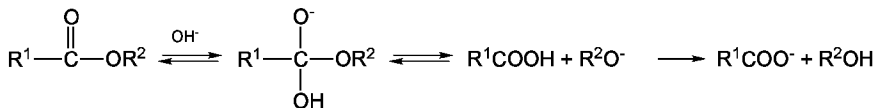
Chemical Properties

The reactions of esters have been reviewed (11–15). Because of the large number of possible acid and alcohol moieties, the chemical properties of esters may differ considerably. Only typical reactions applicable to the majority of esters are described in the following sections.

Hydrolysis. Esters are cleaved (hydrolyzed) into an acid and an alcohol through the action of water. This hydrolysis is catalyzed by acids or bases. The mechanistic aspects of ester hydrolysis have received considerable attention and have been reviewed (16). For most esters only two reaction pathways are important. Both mechanisms involve a tetrahedral intermediate and addition-elimination reactions:*Acid*



Base



Hydrolysis reactions involving tetrahedral intermediates are subject to steric and electronic effects. Electron-withdrawing substituents facilitate, but electron-donating and bulky substituents retard basic hydrolysis. Steric effects in acid-catalyzed hydrolysis are similar to those in base-catalyzed hydrolysis, but electronic effects are much less important in acid-catalyzed reactions. Higher temperatures also accelerate the reaction.

The catalysis of ester hydrolysis by other groups within the ester molecule (intramolecular catalysis) has been extensively studied (17,18). These reactions are important because they simulate catalysis by enzymes. Intramolecular catalysis of esters has been used as a model in drug discovery efforts (19).

Basic Hydrolysis. Throughout most of history, soap was manufactured by boiling an ester with aqueous alkali. In this reaction, known as saponification, the ester is hydrolyzed with a stoichiometric amount of alkali. The irreversible formation of carboxylate anion drives the reaction to completion.

Acidic Hydrolysis. Hydrolysis of esters by use of water and a mineral acid leads to an equilibrium mixture of ester, alcohol, and free carboxylic acid. Complete reaction can only be achieved by removal of alcohol or acid from the equilibrium. Because esters have poor solubility in water, the reaction rate in dilute acids is fairly low. Therefore, emulsifiers such as sulfonated oleic acid or sulfonated aromatic compounds (Twitchell reagent) are added to facilitate the reaction.

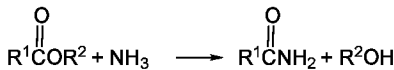
Hydrolysis by Steam. High pressure steam, 4.5–5.0 MPa (650–725 psi), at 250°C in the absence of a catalyst hydrolyzes oils and fats to the fatty acids and glycerol (20). The reaction is commonly carried out continuously in a countercurrent method. The glycerol produced during the reaction is continuously extracted from the equilibrium mixture with water. A yield of 98% can be achieved. Currently, the preferred method to produce soaps is steam hydrolysis of fats followed by alkali neutralization of the fatty acids.

Enzymatic Hydrolysis. Enzymatic hydrolysis has received enormous attention (21–24). The enzymes generally employed are lipases from microorganisms, plants, or mammalian liver. They effect hydrolysis below 40°C. However, this temperature limit can be raised by employing the enzymes from thermophilic bacteria. The enzymes may be used as a crude extract, in purified form or entrapped on a solid support. The great advantage of the enzymatic process is its high chemo- and stereoselectivity. Enzymatic hydrolysis has been used to effect partial hydrolysis of triglycerides, chiral separations of racemic esters, and selective production of specific fatty acids from fats (25). For example, lipase from *Candida cylindracea* was employed to resolve racemic mixtures of R- and S-α-methylarylacetic acid esters to yield S-α-methylarylacetic acids (26).

Transesterification. When esters are heated with alcohols, acids, or other esters in the presence of a catalyst, the alcohol or acid groups are exchanged. This process is called transesterification. It is accelerated by the presence of a small amount of acid or alkali. Three types of transesterification are known: (1) exchange of alcohol groups (alcoholysis), (2) exchange of acid groups (acidolysis), and (3) ester–ester interchange (see ESTERIFICATION). Alcoholysis and acidolysis are important for preparative purposes. All three are equilibrium reactions and proceed to completion if one component is removed from the reaction mixture, eg, by distillation. Dispersed alkali metals, mainly sodium, alkali metal oxides, and tin salts, are suitable catalysts for the transesterification of fats. Recently, organic titanates have also been used (27). Enzymes can be used as asymmetric catalysts in these reactions to prepare optically active alcohols and esters (28).

Transesterification has a number of important commercial uses. Methyl esters of fatty acids are produced from fats and oils. Transesterification is also the basis of recycling technology to break up poly(ethylene terephthalate) [25038-59-9] to monomer for reuse (29) (see RECYCLING, PLASTICS). Because vinyl alcohol does not exist, poly(vinyl alcohol) [9002-89-5] is produced commercially by base-catalyzed alcoholysis of poly(vinyl acetate) [9003-20-7] (see VINYL POLYMERS). An industrial example of acidolysis is the reaction of poly(vinyl acetate) with butyric acid to form poly(vinyl butyrate) [24991-31-9].

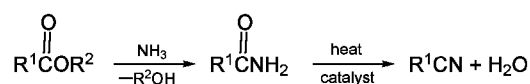
Ammonolysis and Aminolysis. Esters and ammonia react to form amides and alcohols:



This reaction can be carried out in aqueous or alcoholic ammonia. Lower mol wt esters give good yields even at room temperature; higher mol wt esters require higher temperature and pressure.

If primary or secondary amines are used, N-substituted amides are formed. This reaction is called aminolysis. Hydrazines yield the corresponding hydrazides, which can then be treated with nitrous acid to form the azides used in the Curtius rearrangement. Hydroxylamines give hydroxamic acids.

When esters are passed with ammonia over a contact catalyst such as alumina or thoria at 400–500°C, nitriles are obtained via dehydration of the intermediate amides:



Thus fats are converted to the fatty nitriles (30).

Reduction. Esters can be reduced to alcohols by catalytic hydrogenation using molecular hydrogen or by chemical reduction:



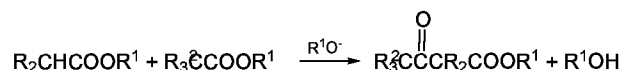
Catalytic Hydrogenation. Esters can be hydrogenated to primary alcohols using a transition-metal catalyst (31–33) such as copper chromite, copper oxide, Raney nickel, nickel–copper–aluminum–rhenium oxides, or related species. The catalyst of choice is copper chromite. Conditions are stringent: 10–30 MPa (1450–4350 psi) at 150–300°C. Halogens and sulfur are catalyst poisons. When the ester is aliphatic and saturated, the reaction is facile and almost quantitative. Catalysts containing Zn or Cd salts have been developed to convert unsaturated fatty esters into unsaturated fatty alcohols (33,34). The reduction of aromatic carboxylic acid esters proceeds beyond the alcohol in some instances. Benzylic C–O hydrogenolysis, eg, benzyl alcohol to toluene, and aromatic ring hydrogenation upon phenol ester reduction are frequent problems. These problems can be minimized by carrying out the reaction at low temperatures with a high ratio of catalyst to ester.

The catalytic hydrogenation of esters is of great commercial importance. It is one of the industrial methods used to produce long-chain fatty alcohols (eg, dodecyl and decyl alcohols) from fatty acid methyl esters (33). The method is also suitable for the conversion of dimethyl 1,4-cyclohexanedicarboxylate [94-60-0] into 1,4-cyclohexanedimethanol [105-08-8], an important intermediate in the manufacturing of polyesters.

Reduction with Metals and Metal Hydrides. Practically any ester can be reduced by Na–C₂H₅OH, Li or Na–NH₃, LiAlH₄, LiBH₄, or NaBH₄ to give alcohols in excellent yield (35,36). Carbon-carbon double bonds are usually preserved using these reducing reagents.

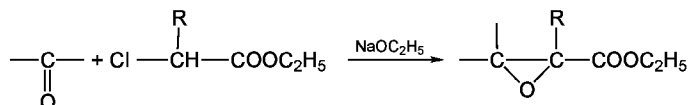
The reduction of esters to aldehydes is carried out with hydrides such as NaH₂Al(OCH₂CH₂OCH₃)₂, (*i*-C₄H₉)₂AlH, NaAlH₄, or LiAlH₄–(C₂H₅)₂NH. The use of BH₃ or LiAlH₄–BF₃·O(C₂H₅)₂ as a reducing reagent converts esters to ethers. Thus, reduction of esters can be manipulated by the judicious selection of metal-containing reducing reagents.

Reaction of Enolate Anions. In the presence of certain bases, eg, sodium alkoxide, an ester having a hydrogen on the α-carbon atom undergoes a wide variety of characteristic enolate reactions. Mechanistically, the base removes a proton from the α-carbon, giving an enolate that then can react with an electrophile. Depending on the final product, the base may be consumed stoichiometrically or may function as a catalyst. For example, the sodium alkoxide used in the Claisen condensation is a catalyst:



The intramolecular Claisen condensation of diesters, or Dieckman reaction, occurs readily to give five- or six-membered rings, and it has been extensively used for cyclopentanone and cyclohexanone derivatives.

Condensations of aldehydes or ketones with α-halo esters give α,β-epoxy esters. This is called the Darzens condensation.



The lithium enolate generated using lithium diisopropylamide [4111-54-0], lithium 2,2,6,6-tetramethylpiperidide [38227-87-1], or lithium hexamethyldisilazide [4039-32-1] is a chemical reagent that reacts with other reactants to give a variety of products (37). In the quest for improved stereospecificity, enolates with different cations such as silicon, aluminum, boron, and zinc have also been used (38). In group transfer polymerization, ketene silyl acetals, eg, (CH₃)₂C=C[OSi(CH₃)₃](OCH₃) are employed as initiators (39).

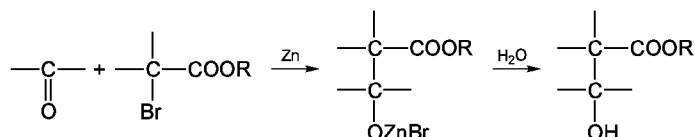
Grignard and Related Reactions. Esters react with alkyl magnesium halides in a two-stage process to give alcohols:



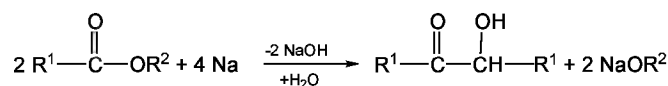
The reaction involves nucleophilic substitution of R³ for OR² and addition of R³MgX to the carbonyl group. With 1,4-dimagnesium compounds, esters are converted to cyclopentanols (40). Lactones react with Grignard reagents and give diols as products.

Many other organometallic compounds also react with carbonyl groups. Lithium alkyls and aryls add to the ester carbonyl group to give either an alcohol or an olefin. Lithium dimethylcuprate has been used to prepare ketones from esters (41). Tebbe's reagent, Cp₂TiCH₂AlCl(CH₃)₂, where Cp = cyclopentadienyl, and other metal carbene complexes can convert the C=O of esters to C=CR₂ (42,43).

α-Halo esters react with aldehydes or ketones in the presence of zinc to form β-hydroxy esters. This is known as the Reformatsky reaction (44).

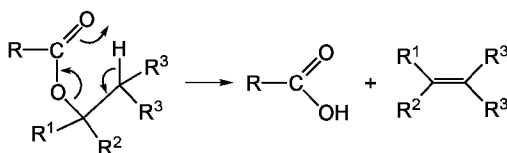


Preparation of Acyloins. When aliphatic esters are allowed to react with metallic sodium, potassium, or sodium–potassium alloy in inert solvents, acyloins (α-hydroxyketones) are formed (45):

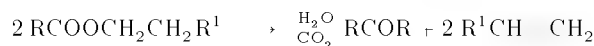


This reaction is used in the synthesis of large-ring compounds.

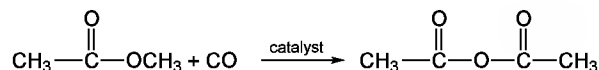
Pyrolysis. The pyrolysis of simple esters of the formula RCOOCR¹R²CHR³₂ to form the free acid and an alkene is a general reaction that is used for producing olefins:



The pyrolysis is generally carried out at 300–500°C over an inert heat-transfer agent such as Pyrex glass or 96% silica glass chips. Esters of tertiary alcohols are pyrolyzed more readily than esters of secondary alcohols, and esters of primary alcohols are the most difficult to pyrolyze. A detailed review on this reaction has been given (46). However, when heated to high temperatures in the presence of metal oxides such as thorium oxide, calcium oxide, manganese chromite, or zinc chromite, esters of primary alcohols give high yields of ketones (47):

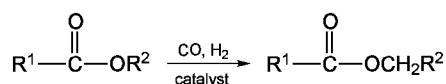


Carbonylation Reaction. The carbonylation of methyl acetate is an important industrial reaction for producing acetic anhydride:

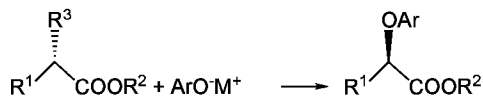


Earlier catalysts were based on cobalt, iron, and nickel. However, recent catalytic systems involve rhodium compounds promoted by methyl iodide and lithium iodide (48,49). Higher mol wt alkyl esters do not show any particular ability to undergo carbonylation to anhydrides.

Ruthenium complexes have been used in the hydrocarbonylation of simple esters to produce the corresponding homologous esters (50). The hydrocarbonylation affects the alkyl moiety rather than the carboxylate group:



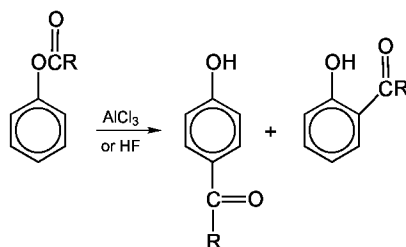
Substitution, Alkylation, and Rearrangement. The reaction of alkaline phenoxides with alkyl *S*-2-(chloro)- or *S*-2-(mesyloxy)propionate gives optically active *R*-2-aryloxyalkanoic acid esters in good chemical and optical yields (>97% ee) (51–53):



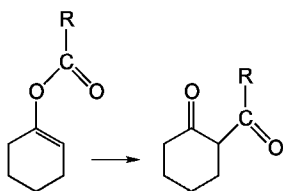
The reaction is utilized in the synthesis of several phenoxy herbicides.

Optically active 2-arylalkanoic acid esters have been prepared by Friedel-Crafts alkylation of arenes with optically active esters, such as methyl *S*-2-(chlorosulfonyl)- or *S*-2-(mesyloxy)propionate, in the presence of aluminum chloride (54,55).

The Fries rearrangement of phenol esters gives a mixture of 2- and 4-acylphenols (56). The reaction is catalyzed by Lewis acids such as aluminum chloride or by Brønsted acids like hydrogen fluoride. This reaction is used in the production of 4-hydroxyacetophenone [99-93-4], a raw material for acetaminophen [103-90-2] (57):



Similarly, enol esters undergo rearrangement to give the corresponding 1,3-diketones. This reaction can be accomplished thermally (500°C) or with a catalyst (58,59):



Occurrence and Preparation Currently, most of the simple esters used commercially are of synthetic origin, although esters occur naturally in large quantities in fats, oils, and waxes. Fats and oils from plants and animals consist mainly of glycerol esters of stearic, palmitic, and oleic acids (60). Natural waxes are esters of monobasic carboxylic acids with higher monohydric and, less commonly, dihydric alcohols. Microorganisms produce a complex array of compounds containing the ester linkage, ranging from simple esters to macrocyclic lactones, such as erythromycin, which are important because of their antibacterial properties.

Essential oils are obtained from fruits and flowers (61,62). Volatile esters of short- and medium-chain carboxylic acids or aromatic carboxylic acids with short- and medium-chain alcohols are primary constituents of essential oils, eg, ethyl acetate in wines, brandy, and in fruits such as pineapple; benzyl acetate in jasmine and gardenia; methyl salicylate in oils of wintergreen and sweet birch. Most of these naturally occurring esters in essential oils have pleasant odors, and either they or their synthetic counterparts are used in the confectionery, beverage, perfume, cosmetic, and soap industries (see OILS, ESSENTIAL).

Recovery of naturally occurring esters is accomplished by steam distillation, extraction, pressing, or by a combination of these processes. Synthetic esters are generally prepared by reaction of an alcohol with an organic acid in the presence of a catalyst such as sulfuric acid, *p*-toluenesulfonic acid, or methanesulfonic acid. Ion-exchange resins of the sulfonic acid type can also be used, and an azeotroping agent such as benzene, toluene, or cyclohexane can be used to remove water and force the reaction to completion (see ESTERIFICATION).

Analysis, Specifications, and Standards

Esters are often characterized by their physical properties. These include boiling point, freezing point, melting point, density, refractive index, residue or ash

content, color, odor, and solubility. An overview of the different analytical methods used on organic esters has been written (63). The most common analytical method is the determination of the saponification value, ie, the number of milligrams of KOH necessary to hydrolyze 1.0 g of ester (64,65). Unsaturated esters, eg, natural fats and oils, are often sufficiently characterized by their iodine value. The iodine value is a measure of the number of double bonds in the molecule.

Standard analytical methods and techniques have been developed for the testing of organic esters and determination of specifications. Some of these specifications and the appropriate ASTM methods of analysis include distillation range (ASTM D1078-86), acidity (ASTM D1613-85), color (ASTM D1209-84), nonvolatile material (ASTM D1353-90), odor (ASTM D1296-84), purity (ASTM D3362-84), and water (ASTM D1364-90) and alcohol content (ASTM D3545-90).

Many chromatographic methods are applicable to organic ester analysis. Liquid chromatography, both normal and reverse phase, is used for all types of esters. Thin-layer and gas–liquid chromatography have been used for analysis of long-chain alkyl esters (waxes) and acrylates (66,67). If enantiomeric resolution is desired, then specialized stationary phases can be used with gas or liquid chromatography (68). Gas chromatographic and gas chromatographic–mass spectral methods of detection are required by the United States Government for monitoring and detection of some organic esters (69).

Esters are usually readily identified by their spectroscopic properties (70). Among these, infrared spectroscopy (ir) is especially useful for identifying the carbonyl of the ester group that has characteristic absorption bands. The C=O absorption is very strong in the ir at 1750–1735 cm^{–1}; in addition, C—O stretching bands are observed in the range of 1100–1300 cm^{–1}. Another spectroscopic method used in identifying organic esters is nuclear magnetic resonance (nmr). The ¹H nmr spectra of esters are characteristic for those protons alpha to the carbonyl group. The peaks for these protons show chemical shifts relative to tetramethylsilane [75-76-3] (TMS) between δ 1.8–2.5 ppm. The peaks for protons alpha to the oxygen function appear between δ 3.3–4.0 ppm relative to TMS. Typical chemical shifts for olefinic protons of α,β-unsaturated esters and enol esters are between 4.5 and 7.5 ppm down field from TMS. The CH proton of formates gives a signal at about 8.0 ppm down field from TMS, and formate esters are therefore easily identified. Another diagnostic method is ¹³C nmr spectroscopy (71,72) which is effective for the detection of ester groups since the ¹³C resonance assignable to the carbonyl carbon of the ester group is observable in the range 160–180 ppm downfield from TMS, and is usually distinguishable from other types of carbonyl carbon atoms. Another useful method for ester determination is mass spectroscopy. The mass spectroscopy of esters (73) has been extensively investigated, and a number of general fragmentation processes have been recognized that may be useful for structure determination.

Stability and Storage

All organic esters are unstable in the presence of acid or base and nucleophiles such as water or alcohols. However, if stored anhydrous, they are stable. Storage vessels can be constructed of steel, aluminum, or other metallic materials, but plastic storage tanks are unsuitable because the highly lipophilic esters can sometimes permeate into the container boundary and soften or even dissolve it. When esters are stored in tanks, a nitrogen blanket and vent are necessary because of potential fire hazards. However, acrylates should be stored in the presence of sufficient oxygen to enable the inhibitor to be effective in preventing polymerization. The esters with high melting points can be stored in paper or wood containers. Proper placarding, packaging, and labeling should always be used before transporting organic esters (74). Sometimes special precautions such as adding inhibitors must be taken before transporting polymerizable esters such as methyl acrylate (75).

The properties of flash point, autoignition temperature, and flammable limit should be considered when an ester is to be handled in any fashion. The flash point is the temperature at which a liquid gives off enough vapor to form an ignitable mixture with air. The flammable limits are the concentrations in air beyond which propagation of flame cannot occur. These limits are usually given as upper and lower concentrations. If the volume percent of the substance is kept above or below these limits, then the mixture will not sustain a flame in oxidizing atmospheres such as air. The autoignition temperature is the temperature required to enable self-sustained combustion of a substance. This information for some selected organic esters is given in Table 2 along with the NFPA ratings of health, flammability, and reactivity. The NFPA ratings are issued on a scale from 1 to 4, the higher value indicating the highest degree of flammability, toxicity, or reactivity.

Table 2. Flammability and Toxicity of Organic Esters

Ester	Auto-ignition temp, °C ^a	Upper flammable limit, vol % in air ^b	Lower flammable limit, vol % in air ^b	NFPA ^c hazard information toxicity data						
				Health	Flammability	Reactivity	Species	Oral LD ₅₀ , g/kg	PEL ^d , ppm	PEL ^d , mg/m ₃
methyl formate	449	23	4.5	2	4	0	rabbit	1.622	100	250
ethyl formate	455	16.0	2.8	2	3	0	rabbit	2.075	100	300
butyl formate	322	8.2	1.7	2	3	0	rabbit	2.66		
methyl acetate	454	16	3.1	1	3	0	rabbit	3.7	200	610
ethyl acetate	426	11.5	2	1	3	0	rabbit	4.97	400	1400
vinyl acetate	402	13.4	2.6	2	3	2	rat	2.92	10	35 ^e
propyl acetate	450	8	1.7(38°C)	1	3	0	rabbit	6.64	200	840
isopropyl acetate	460	8	1.8(38°C)	1	3	0	rabbit	6.95	250	950
butyl acetate	425	7.6	1.7	1	3	0	rat	14	150	710
isobutyl acetate	421	10.5	1.3	1	3	0	rabbit	4.8	150	700
sec-butyl acetate	422	9.8	1.7	1	3	0			200	950
tert-butyl acetate		7.3	1.3						200	950
pentyl acetate	360	1.1	7.5	1	3	0	rat	16.6	100	532 ^e
isoamyl acetate	360	7.5	1(100°C)	1	3	0	rabbit	7.42	100	525
sec-hexyl acetate	266	5	0.9				rat	6.16	50	300
2-ethylhexyl acetate	268	8.14	0.76	2	2	0	rat	>3.2		
ethylene glycol diacetate	482	8.4	1.6	1	1	0	rat	6.86		
2-methoxyethyl acetate	394	8.2	1.7	0	2	0	rat	3.39	25	120
2-ethoxyethyl acetate	380	12.7(135°C)	1.7	2	2	0	rat	5.1	5	27 ^e
2-butoxyethyl acetate	340	8.54(135°C)	0.88(93°C)				mouse	1.6		
2-(2-ethoxyethoxy)ethyl acetate	360	23.5(182°C)	0.98(135°C)				rat	8.69		
2-(2-butoxyethoxy)ethyl acetate	290	24.6	0.76(135°C)				rat	11.9		
benzyl acetate	460	6.1	1(189°C)	1	1	0	rat	2.5		
glyceryl triacetate	433	6.4	1.0(189°C)	1	1	0	mouse	1.5 ^f		
glyceryl tripropionate	421		0.8(186°C)	0	1	0	rat	6.4		
ethyl 3-ethoxypropionate	377	8.7	1.05(88°C)				rat	5		
methyl acrylate	468	25	2.8	2	3	2	rat	3	10	35

ethyl acrylate	372	14	1.4	2	3	2	rabbit	1	25	100
butyl acrylate	292	9.9	1.7	2	2	2	rat	3.7	10	52 ^e
2-ethylhexyl acrylate	252	6.4	0.8	2	2	2	rat	5.6		
methyl methacrylate		8.2	1.7	2	3	2	rabbit	6.55	100	410
methyl butyrate		8.8	1.6	2	3	0	rabbit	3.38		
ethyl butyrate	463	7.7	1.3	0	3	0	rabbit	5.23		
butyl butyrate		6.1	1	2	2	0	rabbit	9.52		
methyl isobutyrate		9	1.6				rat	16		
ethyl isobutyrate		7.8	1.3	0	3	0	mouse	0.8 ^f		
isobutyl isobutyrate	432	7.59	0.96	0	2	0	rat	12.8 ^g		
methyl stearate				0	1	0				10 ^e
ethyl stearate										
butyl stearate	355	4.9	0.3	1	1	0	rat	>32		
dodecyl stearate										
hexadecyl stearate										
dimethyl maleate		10.4	1.6	1	1	0	rat	1.41		
dimethyl oxalate		8.4	1.5	0	2	0	rat	0.4–1.6		
dimethyl adipate							rat	1.81 ^f		
diethyl adipate							rat	>1.6		
di(2-ethylhexyl) adipate	377		0.4(242°C)	0	1	0	rat	9.1		
methyl benzoate		6.7	1.2	0	2	0	rabbit	2.17		
ethyl benzoate	490	6.1	1	1	1	0	rabbit	2.63		
methyl salicylate	454	7.2	1.2	1	1	0	rabbit	2.8		
ethyl salicylate							rat	1.32		
dimethyl phthalate	490	5.8	0.9(180°C)	0	1	0	mouse	7.2 ^f		5
diethyl phthalate	457	5.3	0.7(186°C)	0	1	0	rabbit	1		5 ^e
dibutyl phthalate	402	5.3	0.5(235°C)	0	1	0	rat	8		5 ^e
di(2-ethylhexyl) phthalate	390	5.3	0.3(245°C)				rat	>26		5
dimethyl isophthalate		5.8	1	0	1	0	rat	4.39		
dimethyl terephthalate	518	5.5	0.03	1	1	0	rat	>3.2		
methyl anthranilate				0	1	0	rat	2.91		
benzyl cinnamate							rat	5.53		
dimethyl carbonate				2	3	1	rat	13		
diethyl carbonate		12.4	1.7	2	3	1	mouse	8.5		

^c See Ref. 8.

^a ASTM D286 and D2155.

^b ASTM E6181 (temperature at which limit was determined).

^d Permissible exposure limit; see Ref. 76.

^e Threshold limit value–time-weighted average.

^f Intraperitoneal.

^g LD₁₀₀.

If an organic ester is released, then appropriate action must be taken. The United States Department of Transportation has recommendations for responding to such an event. This includes treating the released material as flammable and poisonous if inhaled or absorbed through the skin. Another recommendation is to be aware that combustion may produce irritating or poisonous gases, thus requiring a positive pressure self-contained breathing apparatus to be worn if exposure is possible. Finally, for fighting fires containing organic esters, dry chemical, CO₂, water spray, or alcohol-resistant foam extinguishing media should be used (77).

Health and Safety Factors

Toxicity. The degree of toxicity of organic esters covers a wide range (78). These toxicities are usually described in terms of threshold limiting values (TLV), or permissible exposure limits (PEL). Both the PEL and the TLV describe the average concentration over an 8-h period to which a worker may be exposed without adverse effects (79). The PEL and TLV data are often interchangeable, although OSHA uses the PEL values (Table 2). The lethal dosages for 50% of the exposed animals, LD₅₀s, are also used as an indicator of the relative toxicity. An accumulation of the LD₅₀ data of organic esters for rabbits, rats, and mice can also be found in Table 2. The LD₅₀s of organic esters for small mammals range between 0.4 and 16 g/kg. The TLVs of organic esters range between 5 and 400 ppm.

When ingested or absorbed, organic esters are likely to be hydrolyzed to the corresponding alcohols and carboxylic acids. Therefore the toxicities of the hydrolysis products should also be considered (80,81). Some organic esters are highly volatile and can act as asphyxiant or narcotic. Also, skin absorption and inhalation are among the hazards associated with esters that are volatile or have good solvent action. Because of the high solubility of fats and oils in organic esters, prolonged or repeated exposure to skin can cause drying and irritation.

Formate esters generally become less toxic as the alcohol moiety increases up to C₄. With this increase in alkyl size, the LD₅₀ (oral, rabbit) increases from 1.62 g/kg for methyl formate to 3.0 g/kg for isoamyl formate [110-45-2]. In comparison, both allyl and vinyl formates are more toxic than their saturated analogues.

Acetates generally do not cause any physiological effects unless high exposure occurs since they are usually converted into or occur naturally as metabolites. However, large enough exposure to acetate esters can cause narcotic effects. The aromatic acetate esters cause death more rapidly than aliphatic acetates with oral LD₅₀s (rat) ranging between 2.5 and 1.6 g/kg for phenyl and benzyl acetate compared to LD₅₀s (rat) of 4.8 to 8.3 g/kg for methyl through propyl acetate. Vinyl acetate gives approximately the same level of toxicity as the other acetate esters and less so when it is polymerized.

Propionates and higher aliphatic esters generally become less toxic as the size of the alkyl carboxylate increases. As an example, the LD₅₀ (rat, oral) for ethyl nonanoate [123-29-5] is greater than 43 g/kg, and the LD₅₀ (rat, oral) for ethyl heptanoate [106-30-9] is 34.6 g/kg.

The acrylate esters are more physiologically hazardous than their saturated homologues. They are usually lachrymators and irritants, and their toxicities decrease with increasing molecular weights. The LD₅₀s of acrylates usually fall between 1 and 5 g/kg for rabbits. Methacrylate esters are generally less toxic than their corresponding acrylates. The decreased physiological activity is believed to result from added steric hindrance of the α-methyl group, but the methacrylates are potent sensitizers.

Among adipates, oxalates, malonates, and succinates, the adipates are the least toxic. An example of this can be seen in the comparison between

di(2-ethylhexyl) adipate, which has an oral LD₅₀ rat of 9.1 g kg, and di(2-ethylhexyl) succinate [117-81-7], which has an oral LD₅₀ rat of 4.3 g kg. The malonates and oxalates are generally more toxic than the adipates. Exposure to diethyl oxalate [95-92-1], the most common oxalate, gives symptoms similar to exposure to oxalic acid [114-62-7], ie, twitching and convulsions. The malonates presumably are less toxic than oxalates because the corresponding malonic acid sodium and calcium salts are much more soluble than calcium oxalates and are thus more easily excreted.

Benzoate esters, like most organic esters, are not very toxic. They are not absorbed through the skin as rapidly as alkyl esters but are more potent physiologically. They are also moderate skin irritants. The oral LD₅₀s (mouse) for methyl- to butylhydroxy benzoates range between 8 and 5 g kg.

The phthalate esters are one of the most widely used classes of organic esters, and fortunately they exhibit low toxicity (82). Because of the ubiquitous nature of phthalates, many investigations have been conducted to determine their toxicities to marine life as well as in mammals (83–85). Generally, phthalates are not absorbed through the skin and are not very potent when inhaled. The phthalates become less toxic as the alcohol group increases in molecular weight. For example, dimethyl phthalate has an oral LD₅₀ (mouse) of 7.2 g kg, whereas di(2-ethylhexyl) phthalate shows an oral LD₅₀ (rat) of greater than 26 g kg.

More information on the toxicities of a range of organic esters is available in the literature (86,87).

Exposure Limits. The Occupational Safety and Health Act (OSHA) of 1990 lists a multitude of acetates, phthalates, formates, and acrylates along with the corresponding permissible exposure limits and threshold limit values (76). The PEL data is listed in Table 2. If there is potential for exposure to an organic ester for which PEL or TLV data has been identified, then an exposure limit lower than that listed is usually selected for working in that environment.

Regulation and Waste

Waste from production of organic esters is usually not a problem since the method of synthesis often involves a carboxylic acid condensation with an alcohol and the only by-product is water. Any organic remnants lost to the process water can usually be biologically degraded. The biochemical oxygen demand (BOD) or chemical oxygen demand (COD) should be measured if biological treatment is used on the process waste from ester production (87). Organic ester vapor emitted in processing usually can be burned.

Extensive federal environmental regulations exist that govern organic esters as well as many other substances (88). These regulations must always be consulted for complete information before using large amounts of organic esters (89). State and local regulations must also be met, which in some cases are more stringent than federal regulations.

Among these federal regulations, the Clean Air Act regulates the amount of an organic ester or other substance that is allowed to be emitted into the atmosphere. Several organic esters are listed as Hazardous Air Pollutants in the Clean Air Act amendments of 1990, and therefore are more tightly regulated (90). If an organic ester is sent to a wastewater treatment facility and subsequently discharged to surface waters, then compliance with the Clean Water Act is required (91). If an organic ester or other substance becomes a solid waste as defined under the Resource Conservation and Recovery Act (RCRA), then specific requirements apply that regulate the treatment, storage, and disposal of that waste. Regulations under RCRA and the Department of Transportation also apply that pertain to the proper labeling, manifesting, and shipping of hazardous wastes (92). The Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) provides a list of hazardous substances, some of which are organic esters (93). The organic esters listed in CERCLA must be properly reported if spilled or otherwise released to the environment in amounts exceeding the reportable quantities specified. Note that spills or releases may need to be reported to state agencies even in amounts that do not exceed the CERCLA reportable quantity. The Superfund Amendments and Reauthorization Act (SARA) of 1986 established regulations requiring facilities to annually report organic esters and other chemicals stored on-site in amounts exceeding reporting thresholds in pure form or as percentages in mixtures. Facilities must also report certain organic esters and other chemicals (listed under SARA 313) that are stored in amounts exceeding reporting thresholds that are released to the environment via air, water, or off-site disposal (94).

Uses

Table 3 lists only those carboxylic acid esters whose 1990 U.S. production, sales quantity, value, and raw materials have been published (95). They are grouped on the basis of their principal use. For a complete list of the organic esters produced and sold in the United States and their manufacturers, the original publication should be consulted. Uses of some specific esters are also given in Table 4.

Table 3. U.S. Production and Sales of Carboxylic Esters, 1990

Material	CAS Registry Number	Production, t	Sales	
			Quantity, t	Value, 10 ³ \$
Plasticizers				
phthalic anhydride esters, total		573,892	572,137	544,485
dibutyl phthalate (including diisobutyl phthalate)	[84-76-2]	7,917	7,714	7,936
diisononyl phthalate	[28553-12-0]	93,575	93,911	78,011
dimethyl phthalate (including dimethyl isophthalate)	[131-11-3]	5,679	5,194	5,830
dioctyl phthalates		140,649	149,805	119,159
all other phthalic anhydride esters		326,072	315,513	333,549
trimellitic acid esters		22,942	28,631	44,524
adipic acid esters, total		87,020	46,683	79,371
di(2-ethylhexyl) adipate	[103-23-1]	24,228	24,587	30,489
diisodecyl adipate	[1330-86-5]	1,494	732	1,211
all others adipic acid esters		61,298	21,364	47,662
complex linear polyesters and polymeric plasticizers		52,904	28,820	59,484
epoxidized esters		47,456	46,518	54,132
butyl oleate	[142-77-8]	805	775	1,149
sebacic acid esters, total		3,139	2,902	14,995
dibutyl sebacate	[109-43-3]	259	268	1,008
all other sebacic acid esters		2,880	2,634	13,987
stearic acid esters, total		5,093	4,850	8,632
isobutyl stearate	[646-13-9]	3,444	3,406	4,338
all other stearic acid esters		1,649	1,444	4,294
Surface-active agents				
carboxylic acid esters, total		166,424	129,375	238,044
anhydrosorbitol esters, total		18,986	14,303	24,106
anhydrosorbitol monolaurate	[1338-39-2]	3,744	2,399	4,594
anhydrosorbitol monooleate	[1338-43-8]	4,442	2,493	4,676
anhydrosorbitol monostearate	[1338-41-6]	8,782	7,919	12,170
all other anhydrosorbitol esters		2,018	1,492	2,666
diethylene glycol esters, total		2,857	1,653	3,701

diethylene glycol monolaurate	[141-20-8]	105	105	172
all other diethylene glycol esters		2,752	1,548	3,529
ethoxylated anhydrosorbitol esters, total		14,292	13,117	29,875
ethoxylated anhydrosorbitol monolaurate	[9005-64-5]	3,315	2,991	7,332
ethoxylated anydrosorbitol monooleate	[9005-65-6]	3,852	3,554	7,486
ethoxylated anhydrosorbitol monostearate	[9005-67-8]	5,117	4,772	10,803
ethoxylated anhydrosorbitol tristearate	[9005-71-4]	273	291	659
all other ethoxylated anhydrosorbitol esters		1,735	1,509	3,595
Surface-active agents				
ethylene glycol distearate	[627-83-8]	1,483	1,434	2,475
ethylene glycol monostearate	[111-60-4]	2,021	1,907	3,554
glycerol esters, total		47,883	37,518	71,108
glycerol dilaurate	[27638-00-2]	233	176	489
glycerol monooleate	[25496-72-4]	4,802	4,926	8,466
glycerol monostearate	[1319-95-5]	3,828	3,869	7,252
all other glycerol esters		39,020	28,547	54,901
natural fats and oils, ethoxylated, total		26,300	18,239	28,530
castor oil, ethoxylated	[61791-12-6]	10,552	9,025	12,366
hydrogenated castor oil, ethoxylated	[61788-85-0]	1,895	1,671	2,233
lanolin, ethoxylated	[61790-81-6]	215	201	529
all other natural fats and oils, ethoxylated		13,638	7,342	13,402
poly(ethylene glycol) esters, total		27,483	21,503	34,975
poly(ethylene glycol) diester of tall oil acids		3,003	832	844
poly(ethylene glycol dilaurate)	[9005-02-1]	724	643	946
poly(ethylene glycol dioleate)	[9005-07-6]	2	610	1,153
poly(ethylene glycol distearate)	[9005-08-7]	920	821	2,726
poly(ethylene glycol monolaurate)	[9004-81-3]	3,456	3,504	5,309
poly(ethylene glycol monooleate)	[9004-96-0]	1,865	1,752	2,312
poly(ethylene glycol monopalmitate)	[9004-94-8]	792		
poly(ethylene glycol monostearate)	[9004-99-3]	3,158	2,928	5,353
poly(ethylene glycol) sesquiester of tall oil acids	[61791-30-8]	865	873	1,652
all other poly(ethylene glycol) esters		10,971	9,540	14,680
poly(glycerol monooleate)	[9007-48-1]	315	265	829
poly(glycerol monostearate)	[37349-34-1]		26	102
1,2-propanediol monostearate	[1323-39-3]	775	199	812
all other carboxylic acid esters		24,029	19,211	37,977
Flavor and perfume materials				
benzyl benzoate	[120-51-4]	245	255	721
phenethyl isobutyrate	[103-48-0]	10		
2-phenethyl phenylacetate	[102-20-5]	24		
cedryl acetate	[77-54-3]	96	30	424
citronellyl acetate	[150-84-5]	41	22	261
citronellyl formate	[105-85-1]	11	5	109
3,7-dimethyl- <i>cis</i> -2,6-octadienol, acetate (neryl acetate)	[141-12-8]	13	12	129
Miscellaneous chemicals				
esters of monohydric alcohols, total		3,113,090	1,481,937	1,354,300
<i>n</i> -butyl acetate	[123-86-4]	114,530	93,242	70,511
butyl acrylate	[141-32-2]	280,129	108,684	129,733
dilauryl-3,3'-thiodipropionate	[123-28-4]	694	704	2,684
distearyl-3,3'-thiodipropionate	[693-36-7]	2,519	2,456	8,168
ethyl acetate (100% basis)	[141-78-6]	123,522	113,668	76,296
ethyl acrylate	[140-88-5]	136,485	66,442	71,791
2-ethylhexyl acrylate	[103-11-7]	53,348	46,300	55,475
fatty acid esters, not included with plasticizers or		5,614	2,955	4,904
surface-active agents, total				
methyl esters of tallow	[61788-61-2]	2,526		
myristyl myristate	[3234-85-3]		88	607
all other fatty acid esters not included with plasticizers or		3,088	2,867	4,297
surface-active agents				
isopropyl acetate	[108-21-4]	20,376	19,299	16,970
methyl methacrylate, monomer	[80-62-6]	536,283		
propyl acetate	[109-60-4]	32,868	30,612	29,968
vinyl acetate, monomer	[108-05-4]	1,206,021	674,970	465,772
all other monohydric alcohol esters		471,306	292,578	357,367
polyhydric alcohol esters, total		145,708	128,357	184,527
2-(2-butoxyethoxy)ethyl acetate	[112-34-5]	4,701	2,957	4,476
2-butoxyethyl acetate	[112-07-2]	8,357	6,492	9,709
glycerides, mixed C14–18 and C16–18, mono-and di-		13,863	13,387	16,993
all other polyhydric alcohol esters		118,787	105,521	153,349

Table 4. Uses of Some Specific Esters

Name and structure	Use
methyl formate, HCOOCH ₃	raw material for production of formamide, dimethylformamide, and formic acid
methyl acetate, CH ₃ COOCH ₃	solvent for cellulose nitrate, cellulose acetate, and many resins and oils; used in the manu-facture of artificial leather; raw material for production of acetic anhydride via carboxylation
ethyl acetate, CH ₃ COOC ₂ H ₅	primarily as a solvent for various resins in pro-TECTIVE coatings; also used extensively in for-mulating printing inks and adhesives; new ap-plications include its uses as a

propyl acetate, CH ₃ COOCH ₂ CH ₂ CH ₃
isopropyl acetate, CH ₃ COOCH(CH ₃) ₂
butyl acetate, CH ₃ COO(CH ₂) ₃ CH ₃
isobutyl acetate, CH ₃ COOCH ₂ CH(CH ₃) ₂
amyl acetates, CH ₃ COOC ₅ H ₁₁
2-ethylhexyl acetate, CH ₃ COOCH ₂ CH(C ₂ H ₅)(CH ₂) ₃ CH ₃
2-butoxyethyl acetate, CH ₃ COOCH ₂ CH ₂ OC ₄ H ₉
2-(2-butoxyethoxy) ethyl acetate, CH ₃ CO(OCH ₂ CH ₂) ₂ OC ₄ H ₉
1-methoxy-2-propyl acetate, ^c CH ₃ CH(OCOCH ₃)CH ₂ OCH ₃
benzyl acetate, CH ₃ COOCH ₂ C ₆ H ₅
ethyl 3-ethoxypropionate, C ₂ H ₅ OCH ₂ CH ₂ COOC ₂ H ₅
isobutyl isobutyrate, (CH ₃) ₂ CHCOOCH ₂ CH(CH ₃) ₂
2,2,4-trimethyl-1,3-pentanediol ^d monoisobutyrate
butyl stearate, CH ₃ (CH ₂) ₁₇ COO(CH ₂) ₃ CH ₃
di(2-ethylhexyl) adipate, [CH ₂ CH ₂ COOCH ₂ CH(C ₂ H ₅)C ₄ H ₉] ₂
benzyl benzoate, C ₆ H ₅ COOCH ₂ C ₆ H ₅
methyl salicylate, 2-OHC ₆ H ₄ COOCH ₃
benzyl salicylate
methyl 4-hydroxybenzoate
methyl cinnamate, C ₆ H ₅ CH=CHCOOCH ₃
2-ethylhexyl 4-methoxycinnamate
dimethyl phthalate
dimethyl terephthalate
di(2-ethylhexyl) phthalate ^e

process solvent in the pharmaceutical industry and as an ex-traction solvent in food processing; as a substi-tute^a for methyl ethyl ketone (MEK) in many applications good solvent for cellulose nitrate, chlorinated rubber, and heat-reactive phenolics; principal use is as a printing ink solvent^b

active solvent for many synthetic resins, such as ethylcellulose, cellulose acetate butyrate, cel-lulose nitrate, some vinyl copolymers, poly-styrene, and methacrylate resins; as a solvent for printing ink; like propyl acetate, it can also be used in the recovery of acetic acid from dilute aqueous solutions

excellent solvent for inks and lacquers because of its high blush resistance and evaporation rate; widely used as solvent in paints, thinner, video tape binders, and extraction of pharma-ceuticals; also used as a perfume ingredient and as a component in synthetic flavors such as apricot, banana, butter, pear, quince, pine-apple, grenadine, butterscotch, and raspberry; also a cleaning solvent for silicon wafers

resembles butyl acetate and methyl isobutyl ketone (4-methyl-2-pentanone) and can be used interchangeably for these solvents in many formulations; also a component in synthetic flavors of apple, apricot, banana, butter, mira-belle plum, pineapple, rum, and strawberry

amyl acetate and mixed amyl acetates (a mix-ture of normal, secondary, and isoamyl ace-tates) are used as lacquer solvents, as extrac-tants in penicillin manufacture, and in the production of photographic film, leather pol-ishes, dry-cleaning preparations, and flavoring agents; mixed *sec*-amyl acetates are used as solvents for cellulose compounds and in the production of leather finishes, textile sizes, and printing compounds; isoamyl acetates are used as solvents and in flavorings and perfumes

high boiling retarder solvent with limited water solubility used to promote flow of and retard blushing in lacquers, emulsions, and silk-screen inks, and as a flow-control agent in baking enamels; also used as a dispersant for vinyl organosols, and as a coalescing aid for latex paints

slow-evaporating glycol ether ester useful as a coalescing aid in poly(vinyl acetate) emulsion system; also used as a retarder solvent in lacquers, enamels, and printing inks solvent in printing inks and high bake enamels; also used as a coalescing aid in latex paints, in silk-screen inks, and as a component in poly-styrene coatings for decals

solvent in inks, ink remover, paints, automotive coatings, and photoresist; also a substitute for 2-ethoxyethyl acetate in many applications

component of the extract of gardenia, hyacinth, and ylang-ylang, and the main component of extract of jasmine; most benzyl acetate is used in soap odors, but it is also popular for other perfumes and is used to a minor extent in flavors

linear ether ester with excellent solvent proper-ties for many of the polymers and resins used in coating industry; provides lower solution viscosity than many other retarder solvents of similar evaporation rate, and it can be a re-placement for 2-ethoxyethyl acetate

a retarder solvent in wood lacquers, automotive coatings, metal coatings, and a variety of thin-ner blends; also used in high solids coatings because of its low surface tension, which im-proves surface characteristics; its distinct odor and flavor make it an interesting material for the formulation of perfumes, and as a bulk component of flavor essences

widely used as a coalescing aid in latex paints, effective with a broad range of latex emulsion systems; retarder solvent for high solid coat-ings, and a sweetener in letterpress and litho-graphic inks to improve solvent activity of ink's solvent system used for compounding lubricating oils, as a lubricant for the textile and molding trade, in special lacquers, and as a waterproofing agent; in the cosmetic and pharmaceutical fields, it is used in vanishing creams, oint-ments, rouges, lipsticks, and nail polishes; its oily characteristics have made it of particular value in polishes and coatings that are to be polished

plasticizer to impart low temperature flexibility to PVC formulations, particularly in vinyl meat-wrapping film

used in perfumery as a fixative, as a solvent for synthetic musks, and in confectionery and chewing gum flavors; also used in medicine and cosmetics and as plasticizer, insect repel-lent, and dye carrier

main component of wintergreen oil and occurs in small quantities in other essential oils and fruit; used primarily for the relief of muscular aches, articular rheumatism, and neuralgia; as a flavor and fragrance agent, it is used in confectionery, dentifrices, cosmetics, and in perfumes; also used as a dye carrier and uv light stabilizer in acrylic resins

widely used in soap and cosmetic industry as fragrance; also effective in absorbing uv light, and can be used in protective sunscreen lotions

broad spectrum of antimicrobial activity, low levels of toxicity, excellent stability and inert-ness; used as preservative in cosmetic formu-lations, general-purpose cleaners, disinfec-tants, and mouth wash and contact lens cleaning solutions; also used as food additive and pharmaceutical preservative

fragrance in soaps, perfumes, and confectioneries

absorbs uv rays effectively; thus about 75° % of all sunscreen formulations use it; usually nonal-lergenic and nonstaining

raw material for polyesters; also used as plasti-cizer, mosquito repellent, dye carrier, and in hair sprays

raw material for polyesters such as poly(ethylene terephthalate), poly(butylene terephthalate), and unsaturated polyester

plasticizer; also used as an insulating fluid in electrical transformers and uppressure-sensitive printing

- ^a Ethyl acetate (exempt solvent) is much less toxic than MEK.
- ^b Compared with ethyl acetate and isopropyl acetate, propyl acetate has slow evaporation rate and good solvent power which promote improved flow and leveling characteristics in a variety of coating formulations.
- ^c Propylene glycol methyl ether acetate.
- ^d $(\text{CH}_3)_2\text{CHCOOCH}_2\text{C}(\text{CH}_3)_2\text{CH}(\text{OH})\text{CH}(\text{CH}_3)_2$.
- ^e Dioctyl phthalate (DOP).

Solvents. Lower esters are extensively used as solvents in coatings (eg, paints and top coats on automobiles), inks, and adhesives, and in processing other substances (96). They readily dissolve resins or their precursors to become vehicles for application. Because these solvent esters are not on the list of 189 hazardous air pollutants regulated by Section 112 of the Clean Air Act of 1990, they will not face the decline in use in the short term that methyl ethyl ketone and methyl isobutyl ketone will. However, in the long term, environmental concerns enforced by regulations to reduce the amount of volatile organic components (VOCs) in air will gradually decrease the usage of esters in solvent applications.

Plasticizers. Plasticizers are materials that soften and flexibilize inherently rigid, and even brittle polymers. Organic esters are widely used as plasticizers in polymers (97,98). These esters include the benzoates, phthalates, terephthalates, and trimellitates, and aliphatic dibasic acid esters. For example, triethylene glycol bis(2-ethylbutyrate) [95-08-9] is a plasticizer for poly(vinyl butyral) [63148-65-2], which is used in laminated safety glass (see VINYL POLYMERS, POLY(VINYL ACETALS)). Di(2-ethylhexyl)phthalate [117-81-7] (DOP) is a preeminent plasticizer. Variation of acid and or alcohol component(s) modifies the efficacy of the resultant ester as a plasticizer. In phthalate plasticizers, molecular sizes of the alcohol moiety can be varied from methyl to tridecyl to control permanence, compatibility, and efficiency; branched (eg, 2-ethylhexyl, isodecyl) for rapid absorption and fusion; linear (C6–C11) for low temperature flexibility and low volatility; and aromatic (benzyl) for solvating. Terephthalates are recognized for their migration resistance, and trimellitates for their low volatility in plasticizer applications.

Resins, Plastics, and Coatings. Unsaturated and difunctional esters are important monomers for the manufacture of many polymers in commercial use. For example, free-radical polymerization of vinyl acetate and methyl methacrylate produces poly(vinyl acetate) [9003-20-7] (PVAc) and poly(methyl methacrylate) [9011-14-7], respectively. Applications of PVAc include latex paint, paper manufacturing, coating for paper board, and adhesives for packaging and labeling (see VINYL POLYMERS, POLY(VINYL ACETATE)). Poly(methyl methacrylate) is used for glazing, lighting fixtures, optical fibers, and surface coatings (see METHACRYLIC POLYMERS). Another example is dimethyl terephthalate (DMT) which reacts with ethylene glycol to yield poly(ethylene terephthalate) [25038-59-9] (PET). PET is used in fibers, films, and bottles (see POLYESTERS). Liquid crystal polymers (LCPs) are a class of thermoplastic polyesters with aromatic carbon backbones. Amoco's Xydar resins are based on terephthalic acid, 4,4'-dihydroxybiphenyl [92-88-6], and 4-hydroxybenzoic acid [99-96-7]. Hoechst Celanese's Vectra resins are based on 4-hydroxybenzoic acid and 6-hydroxy-2-naphthoic acid (see ENGINEERING PLASTICS). LCPs have found application in aviation, electronics (connectors, sockets, chip carriers), automotive underhood parts, and chemical processing. Copolymerization of ethylene with unsaturated esters such as vinyl acetate, methyl acrylate, ethyl acrylate, or butyl acrylate yields polyolefins with special properties. Unsaturated polyesters, produced by condensation of unsaturated dibasic acids (eg, maleic anhydride), and glycols (eg, propylene glycol), are used as thermosets when combined with a cross-linking agent (eg, styrene) in the presence of a free-radical initiator and a promoter. Their applications include boat, automotive exterior parts, cultured marbles, bowling balls, polymeric concrete, and coatings. Polyester polyols are used in polyurethanes (see URETHANE POLYMERS).

Poly(3-hydroxybutyrate–3-hydroxyvalerate) [80181-31-3] resin, produced from a bacterium during a sugar fermentation process, has been reported to be biodegradable, and its target markets include "flushables" such as feminine hygiene products and disposable diapers (99).

Lubricants. Monohydric alcohol esters of dibasic acids and polyol esters of monobasic acids are synthetic lubricants (100). They are generally prepared from the following alcohols and acids: (1) C_8 – C_{13} monohydric alcohols such as 2-ethylhexyl, isooctyl, isodecyl, and isotridecyl alcohols; (2) polymethylol compounds such as trimethylolpropane, pentaerythritol, and dipentaerythritol; (3) C6–C10 monobasic acids such as heptanoic and nonanoic acids; and (4) C6–C10 dibasic acids such as adipic, azelaic and sebacic acids, and phthalic anhydride. These esters are mainly used as base oils in high performance lubricants for automotive (eg, engines), aviation (eg, gas turbines), and machinery (eg, gear, chain, and air compressor) industries. Compared with petroleum oils, ester lubricants exhibit lower pour point, higher thermal and oxidation stability, high viscosity index, lower volatility, and better lubricity. For example, bis(2-ethylhexyl) sebacate [122-62-3] is widely used as base oil for lubricating turbojet engines. Polyol esters are used as textile lubricants because of reduced carbon deposit formation.

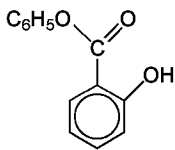
Perfumes, Flavors, Cosmetics, and Soap. Many naturally occurring esters in essential oils and some synthetic esters are important fragrance and flavor compounds (61,62). They are used in perfumes, flavors, cosmetics, soaps, detergents, and air fresheners. Benzyl, butyl, ethyl, methyl, and phenyl esters of benzoic acid are used as flavors, perfumes, and food preservatives. Glyceryl 4-aminobenzoate [136-44-7] and 2-ethylhexyl 4-dimethylaminobenzoate [21245-02-3] are used in cosmetic sunscreen preparations. Alkyl esters of 4-hydroxybenzoic acid, called parabens, have been used under various names for fungus infections of the skin, and as preservatives in lotions and creams (101). Soap and cosmetic fragrances use large amounts of amyl and benzyl salicylate. Benzyl salicylate [118-58-1] is also used in deodorant sprays. 2-Ethylhexyl salicylate [118-60-5] and 2-ethylhexyl 4-methoxycinnamate [5466-77-3] are used in sunscreen formulations (102).

Benzyl-diethyl-[(2,6-xylylcarbamoyl)methyl]ammonium benzoate (denatonium benzoate [3734-33-6], Bitrex) is an extremely bitter tasting, nonirritating, and nonmutagenic compound that has been widely used in many household products such as detergents, nail polish removers, and cleaning agents, to prevent childhood poisoning. It is also used as an alcohol denaturant.

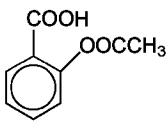
Organic esters in the form of fats and oils from tallow and plants such as soybean, cottonseed, linseed, and castor bean are important raw materials for soap, paints, and food industries.

Surface-Active Agents. Polyol (eg, glycerol, sorbitol, sucrose, and propylene glycol) or poly(ethylene oxide) esters of long-chain fatty acids are nonionic surfactants (qv) used in foods, pharmaceuticals, cosmetics, textiles, cleaning compounds, and many other applications (103,104). Those that are most widely used are included in Table 3.

Medicinals. Many esters are used as pharmaceuticals (105,106). Of these, benzocaine, ethyl 4-aminobenzoate [94-09-7] is a topical anesthetic. Phenyl salicylate [118-55-8] (1) has antipyretic, antirheumatic, and antiparasitic properties.



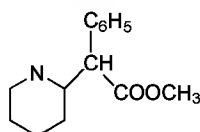
(1)



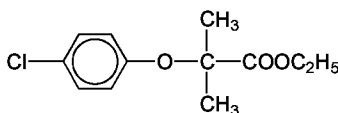
(2)

Some simple benzoates are also used as antiseptic agents. Salicylic acid esters are used as antibacterial agents and pain relievers. Analgesic balms, creams,

sprays, and nasal inhalers usually contain various combinations of either methyl or menthyl salicylate and menthol. In general, esterification of a physiologically active alcohol or phenol with aliphatic carboxylic acid or an acid with alcohol detoxifies it by decreasing the concentration of active compound present. The active compound is released gradually in the body by hydrolysis of the ester (107). Examples include aspirin [50-78-2] (2), a common analgesic; methyl phenidate [113-45-1] (3), a central nervous system stimulant; and clofibrate [637-07-0] (4), a antihyperlipoproteinemic.

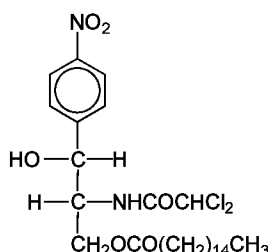


(3)



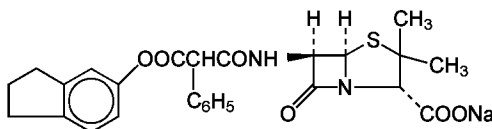
(4)

In recent years, many parent drugs have been converted to esters to generate so-called prodrugs in order to overcome some undesirable property such as bitter taste, poor absorption, poor solubility, and irritation at site of injection. For example, antibiotics such as chloramphenicol [56-75-7] and clindamycin [18323-44-9] have been derivatized as their palmitate esters in order to minimize their bitter taste.

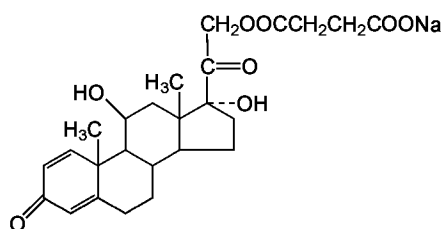


chloramphenicol palmitate

In order to improve the poor oral absorption of carbenicillin [4697-36-3], a lipophilic indanyl ester has been formulated, Geocillin [35531-88-5] (5). Prednisolone [50-24-8], a steroid, is derivatized to its C-21 hemisuccinate sodium salt (6) to make it extremely water-soluble (108).



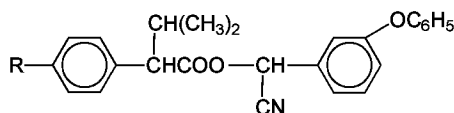
(5)



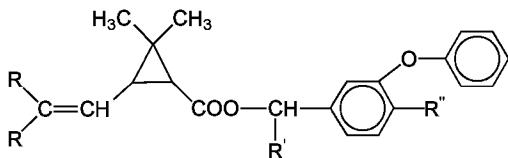
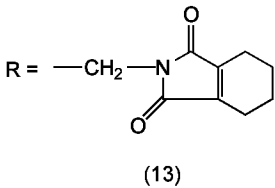
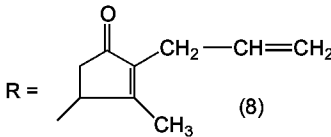
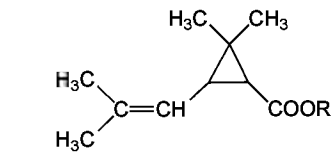
(6)

Herbicides and Pesticides. Several esters are used as herbicides and pesticides (109–111). Many halogenated benzoic acid esters are phytotoxic and are used as herbicides. Dimethyl tetrachloroterephthalate [1861-32-1] and diisopropyl 5-nitroisophthalate [10552-74-6] are used as herbicide and fungicide, respectively. The phenoxy herbicides are primarily propyl, butyl, and isooctyl esters of 2,4-dichlorophenoxyacetic acid [94-75-7], 4-chloro-2-methylphenoxyacetic acid [93-65-2], and methyl, ethyl, or butyl esters of 2-(4-hydroxyphenoxy)propionic acid [67648-61-7]. Because of their low toxicity, high selectivity, and relatively short life in the soil, phenoxy herbicides are widely used. They are used for controlling weeds in a large number of grass crops, ie, corn, small grains, sorghum, rice, sugarcane, pasture, range land, and turf.

Pyrethroids are synthetic esters produced to imitate or improve the activity of biological principles of the pyrethrum plant. They are powerful contact insecticides causing rapid knockdown of treated insects. The pyrethroids are extensively used in controlling insect pests on fruit trees, vegetables, and other field crops; in space sprays and contact sprays to kill insects infesting homes, industrial locations, and nonfood processing areas; and in protection of warehoused food. These compounds include fenvalerate [51630-58-1], (7), R = Cl; flucythrinate [70124-77-5], (7) R = CHF₂O; allethrin [584-79-2] (8); cyfluthrin [68359-37-5] (9); cypermethrin [52315-07-8] (10); deltamethrin [52918-63-5] (11); permethrin [52645-53-1] (12); and tetramethrin [7696-12-0] (13).



(7)



(9) R = Cl; R' = CN; R'' = F

(10) R = Cl; R' = CN; R'' = H

(11) R = Br; R' = CN; R'' = H

(12) R = Cl; R' = H; R'' = H

Miscellaneous Uses. Since esters can be made and hydrolyzed with ease, they are used as protecting groups for hydroxyl and carboxylic acid groups (112). Acetates and benzoates are widely used in carbohydrate, steroid, and nucleoside chemistry, and their cleavage is based on hydrolysis with base, ammonolysis, or methanolysis. Of great importance in peptide chemistry are the *t*-butyl, benzyl, and substituted benzyl esters (113,114). In recent years, esters as protecting groups have played an increasing role in modulating efficacy and bioavailability of pharmaceuticals.

Esters such as benzoates and phthalates are also used in the preparation of high activity catalysts for olefin polymerization. They appear to function as electron donors in the catalyst complex, and play a significant role in catalyst performance (115).

BIBLIOGRAPHY

"Esters, Organic" in *ECT* 1st ed., Vol. 5, pp. 824–950, by T. E. Jordan, Publicker Industries, Inc.; in *ECT* 2nd ed., Vol. 8, pp. 365–383, by E. U. Elam, Tennessee Eastman Co.; in *ECT* 3rd ed., Vol. 9, pp. 311–337, by E. U. Elam, Tennessee Eastman Co.

- W. Kantlehner in B. M. Trost and I. Fleming, eds., *Comprehensive Organic Synthesis: Selectivity, Strategy and Efficiency in Modern Organic Chemistry*, Vol. 6, Pergamon Press, Inc., Elmsford, N.Y., 1991, pp. 485–599.
- J. Voss in S. Patai, ed., *The Chemistry of Acid Derivatives*, Suppl. B, Part 2, John Wiley & Sons, Inc., New York, 1979, Chapt. 18.
- J. Voss in Ref. 1, pp. 435–460.
- J. D. Fletcher, O. C. Dermer, and R. B. Fox, *Nomenclature of Organic Compounds*, ACS Advances in Chemistry Series 126, American Chemical Society, Washington D.C., 1974, pp. 137–145.
- IUPAC Commission on Nomenclature of Organic Chemistry: Nomenclature of Organic Chemistry*, 4th ed., Pergamon Press, Inc., Elmsford, N.Y., 1979.
- E. K. Euranto in S. Patai, ed., *The Chemistry of Carboxylic Acids and Esters*, Wiley-Interscience, New York, 1969.
- J. A. Riddick, W. B. Bunger, and T. K. Sakano, *Organic Solvents*, 4th ed., John Wiley & Sons, Inc., New York, 1986.
- Fire Protection Guide to Hazardous Materials*, 10th ed., National Fire Protection Association, Quincy, Mass., 1991.
- D. R. Lide, ed., *CRC Handbook of Chemistry and Physics*, 71st ed., CRC Press, Inc., Boca Raton, Fla., 1990.
- J. A. Dean, ed., *Lange's Handbook of Chemistry*, 13th ed., McGraw-Hill Book Co., Inc., New York, 1985.
- Ref. 6, Chapt. 11.
- I. O. Sutherland in D. Barton and W. D. Ollis, eds., *Comprehensive Organic Chemistry*, Vol. 2, Pergamon Press, Inc., Elmsford, N.Y., 1979, pp. 913–937.
- B. M. Trost and I. Fleming, eds., *Comprehensive Organic Synthesis: Selectivity, Strategy and Efficiency in Modern Organic Chemistry*, Vol. 1–9, Pergamon Press, Inc., Elmsford, N.Y., 1991.
- R. C. Larock, *Comprehensive Organic Transformations*, VCH Publishers, Inc., New York, 1989.
- J. March, *Advanced Organic Chemistry*, 4th ed., John Wiley & Sons, Inc., New York, 1992.
- C. K. Ingold, *Structure and Mechanism in Organic Chemistry*, 2nd ed., Bell, London, 1969, Chapt. 15.
- T. C. Bruice and S. J. Benkovic, *Bio-Organic Mechanisms*, Vol. 1, Benjamin, New York, 1966.
- A. J. Kirby and A. R. Fersht, *Prog. Bio-Org. Chem.* **1**, 1 (1971).
- W. Sneader, *Drug Discovery: The Evolution of Modern Medicines*, John Wiley & Sons, Inc., New York, 1986.
- N. O. V. Sonntag in R. W. Johnson and E. Fritz, eds., *Fatty Acids in Industry*, Marcel Dekker, Inc., New York, 1989, pp. 23–72.
- J. B. Jones, *Tetrahedron* **42**, 3351 (1986); J. B. Jones in J. D. Morrison, ed., *Asymmetric Synthesis*, Vol. 5, Academic Press, Inc., New York, 1985, p. 309.
- M. Ohno and M. Otsuka in A. S. Kende and co-workers, eds., *Organic Reactions*, Vol. 37, John Wiley & Sons, Inc., New York, 1989, pp. 1–55.
- H. G. Davies and co-workers, *Biotransformations in Preparative Organic Chemistry*, Academic Press, Inc., London, 1989.
- F. X. McConville, J. L. Lopaz, and S. A. Wald in D. A. Abramowicz, ed., *Biocatalysis*, Van Nostrand Reinhold Co., New York, 1990, pp. 167–177.
- A. R. Macrae, *J. Am. Oil Chem. Soc.* **60**, 243A (1983).
- Eur. Pat. Appl. 227,078 A1** (Jan. 7, 1987), C. J. Sih (to Wisconsin Alumini Research Foundation).
- Tygon Organic Titanates*, DuPont Bulletin D5258, E. I. Du Pont de Nemours & Co., Inc., Wilmington, Del., 1972, 5M. Rev. 377.
- A. M. Klibanov, *Acc. Chem. Res.* **23**, 114 (1990).
- R. D. Leaversuch, *Mod. Plast.* 40–43, (July, 1991).
- S. H. Shapiro in E. S. Patterson, ed., *Fatty Acids and Their Industrial Applications*, Marcel Dekker, Inc., New York, 1968, Chapt. 5.

31. P. N. Rylander, *Hydrogenation Methods*, Academic Press, Inc., New York, 1985; P. N. Rylander, *Catalytic Hydrogenation in Organic Syntheses*, Academic Press, Inc., New York, 1979.
32. H. Kropf in E. Müller, ed., *Methoden der Organischen Chemie*, Houben-Weyl, Vol. VI 1b, Georg Thieme Verlag, Stuttgart, 1984, Chapt. 8.
33. U. R. Kreutzer, *J. Am. Oil Chem. Soc.* **61**, 343 (1984).
34. P. N. Rylander in J. R. Anderson and M. Boudart, eds., *Catalysis*, Vol. 4, Springer-Verlag, Berlin, 1983, pp. 1–38.
35. A. Hajos, *Complex Hydrides and Related Reducing Agents in Organic Synthesis*, Elsevier Science Publishing Co., New York, 1979.
36. J. Malek in A. S. Kende and co-workers, eds., *Organic Reactions*, Vol. 36, John Wiley & Sons, Inc., New York, 1988, pp. 249–590.
37. C. H. Heathcock in E. Bunel and T. Durst, eds., *Comprehensive Carbanion Chemistry*, Elsevier, Amsterdam, 1984, Part B, Chapt. 4.
38. H. B. Mekelburger and C. S. Wilcox in Ref. 13, Vol. 2, pp. 99–131.
39. D. Y. Sogah and co-workers, *Macromolecules* **20**, 1473 (1987).
40. P. Canonne and M. Bernatchez, *J. Org. Chem.* **52**, 4025 (1987).
41. K. Fuji and co-workers, *Tetrahedron Lett.* **32**, 7277 (1991).
42. S. H. Pine and co-workers, *J. Org. Chem.* **50**, 1212 (1985).
43. R. R. Schrock, *J. Am. Chem. Soc.* **98**, 5399 (1976).
44. A. Furstner, *Synthesis*, 571 (1989).
45. J. J. Bloomfield, D. C. Owsley, and J. M. Nelke in W. G. Dauben and co-workers, eds., *Organic Reactions*, Vol. 23, John Wiley & Sons, Inc., New York, 1976, pp. 259–403.
46. R. Taylor in Ref. 2, Part 1, pp. 859–914.
47. P. Sabatier and E. E. Reid, *Catalysis Then and Now*, part 2, Franklin Publishing Co., Englewood, N.J., 1965, pp. 308–314.
48. S. W. Polichnowski, *J. Chem. Ed.* **63**, 206 (1986).
49. V. H. Agredan, D. M. Pond, and J. R. Zoeller, *Chemtech*, 172 (Mar. 1992).
50. G. Jenner, *Appl. Catal.* **50**, 99 (1989).
51. **U.S. Pat. 4,173,709** (Nov. 6, 1979), J. Metivier and M. Sauli (to Rhone-Poulenc SA).
52. **U.S. Pat. 4,537,984** (Aug. 27, 1985), I. Hashiba, T. Nishikimi, and S. Tsuchiya (to Nissan Chemical Industries Ltd.).
53. **Jpn. Pat. Appl. 62,178,543** (Aug. 5, 1987), K. Sato and H. Miyoshi (to Daicel Chemical Industry).
54. **U.S. Pat. 4,707,549** (Nov. 17, 1987), G. Fritz, M. Eggersdorfer, and H. Siegel (to BASF AG).
55. O. Piccolo and co-workers, *J. Org. Chem.* **50**, 3945 (1985).
56. H. J. Shine, *Aromatic Rearrangements*, Elsevier Science Publishing Co., New York, 1967, Chapt. 2.
57. **U.S. Pat. 4,560,789** (Dec. 24, 1985), K. G. Davenport and C. B. Hilton (to Hoechst Celanese Corp.).
58. F. G. Young and co-workers, *J. Am. Chem. Soc.* **72**, 3635 (1950).
59. A. A. Akhrem and co-workers, *Synthesis*, 925 (1978).
60. H. Baumann and co-workers, *Angew. Chem. Int. Ed. Engl.* **27**, 42 (1988).
61. K. Bauer, D. Garbe, and H. Surburg, *Common Fragrance and Flavor Materials*, 2nd rev. ed., VCH Publishers, New York, 1990.
62. O. Secondini, *Handbook of Perfumes and Flavors*, Chemical Publishing Co., Inc., New York, 1990.
63. T. S. Ma in Ref. 6, pp. 871–921.
64. *Annual Book of ASTM Standards, D1386-83*, American Society for Testing and Materials, Philadelphia, Pa., 1991.
65. E. Weisenberger, *Mikrochim. Acta* **30**, 241 (1942).
66. A. P. Tulloch, *J. Chromatog. Sci.* **13**, 403 (1975).
67. A. Horna, H. Pechava, and J. Churacek, *J. Chromatog.* **330**(2), 439 (1985).
68. E. Fracotte and D. Lohmann, *Helv. Chim. Acta* **114**, 647 (1989).
69. *Code of Federal Regulations*, Vol. 40, Part 136, Office of the Federal Register National Archives and Records Admin., Washington, D.C., 1990.
70. R. Silverstein, G. C. Bassler, and T. C. Morrill, *Spectrometric Determination of Organic Compounds*, 5th ed., John Wiley & Sons, Inc., New York, 1991.
71. J. B. Stothers, *Carbon-13 NMR Spectroscopy*, Academic Press, Inc., New York, 1972.
72. G. C. Levy and G. L. Nelson, *Carbon-13 Nuclear Magnetic Resonance for Organic Chemists*, John Wiley & Sons, Inc., New York, 1972.
73. H. Budzikiewicz, C. Djerassi, and D. H. Williams, *Mass Spectrometry of Organic Compounds*, Holden Day, San Francisco, 1967, Chapt. 4.
74. Ref. 69, part 172.
75. *Acrylate Esters, A Guide to Safety and Handling*, compiled by Badische Corp., Celanese Chemical Co. Inc., Rohm and Haas Co., and Union Carbide Corp., 1984.
76. *Code of Federal Regulations*, Vol. 29, part 1910, Office of the Federal Register National Archives and Records Administration, Washington, D.C., 1990.
77. *Emergency Response Guide Book*, United States Department of Transportation, Washington, D.C., 1990.
78. *Registry of Toxic Effects of Chemical Substances*, Vol. 1–4, United States Department of Health and Human Services, U.S. Government Printing Office, Washington, D.C., 1988.
79. American Conference of Governmental Industrial Hygienists, *Threshold Limit Values*, Cincinnati, Ohio, 1991.
80. F. Clayton and G. Clayton, *Patty's Industrial Hygiene and Toxicology* Vol. 1, 4th ed., Part B, John Wiley & Sons, Inc., New York, 1991.
81. N. I. Sax and R. J. Lewis, Sr., *Dangerous Properties of Industrial Materials*, 7th ed., Van Nostrand Reinhold Co., New York, 1989.
82. R. J. Jaeger and R. J. Rubin, *Science* **170**, 460 (1970).
83. *In-Depth Studies on Health and Environmental Impacts of Selected Water Pollutants*, report on contract number 68-01-4646, U.S. Environmental Protection Agency, Washington, D.C., 1978.
84. F. L. Mayer, Jr., and H. O. Sanders, *Environ. Health Perspec.* **3**, 153 (1973).
85. D. Peakall, *Residue Rev.* **54**, 1 (1975).
86. K. Gijtsu and S. Kenkyusho, *Toxic and Hazardous Industrial Chemicals Safety Manual*, The International Technical Information Institute, Tokyo, 1988.
87. K. Verschuere, *Handbook of Environmental Data on Organic Chemicals*, 2nd ed., Van Nostrand Reinhold Co., New York, 1983.
88. *Chemical Regulatory Crossreference*, J. J. Keller & Associates Inc., Neenah, Wis., 1991.
89. T. Wagner, *The Complete Handbook of Hazardous Waste Regulation*, Perry-Wagner Publishing Co., Inc., 1988.
90. *U.S. Clean Air Act Amendments*, Public Law 101-549, Title 3, Hazardous Air Pollutants, U.S. Government Printing Office, Washington, D.C., 1990.
91. Ref. 69, part 117.
92. Ref. 69, part 260–281.
93. Ref. 69, part 302.
94. Ref. 69, part 372.
95. "Synthetic Organic Chemicals, United States Production and Sales, 1990" in *U.S. International Trade Commission Publication 2470*, U.S. International Trade Commission, Washington, D.C., 1991.
96. E. W. Fick, *Industrial Solvents Handbook*, 4th ed., Noyes Data Corp., Park Ridge, N.J., 1991.
97. L. Meier in R. Gachter and H. Muller, eds., *Plastics Additives Handbook*, 3rd ed., Hanser Publishers, Munich, 1990.
98. D. Dieckmann in *Modern Plastics Encyclopedia '92*, McGraw-Hill Book Co., Inc., 1991, pp. 184–187.
99. *Mod. Plast.*, 29 (June 1990).

100. H. F. Reid in R. W. Johnson and E. Fritz, eds., *Fatty Acids in Industry*, Marcel Dekker, Inc., New York, 1989, pp. 377–405.

101. J. R. Sabourin, *Drug & Cosmetic Industry*, **147**, 24 (1990).

102. M. N. O'Donoghue, *Dermatol. Clin.* **9**, 99 (1991).

103. R. T. McIntyre and R. W. Johnson and E. Fritz, eds., *Fatty Acids in Industry*, Marcel Dekker, Inc., New York, 1989, pp. 351–375.

104. R. A. Reck in R. W. Johnson and E. Fritz, eds., *Fatty Acids in Industry*, Marcel Dekker, Inc., New York, 1989, pp. 201–215.

105. M. Sittig, *Pharmaceutical Manufacturing Encyclopedia*, Vol. 1–2, 2nd ed., Noyes Publications, Park Ridge, N.J., 1988.

106. S. Budavari and co-workers, eds., *The Merck Index*, 11th ed., Merck & Co., Inc., Rahway, N.J., 1989.

107. S. H. Yalkowsky and W. Morozowich in E. J. Ariens, ed., *Drug Design*, Vol. 9, Academic Press, Inc., New York, 1980, Chapt. 3.

108. L. K. Low and N. C. Castagnoli, Jr. in R. F. Doerge, ed., *Wilson and Gisvold's Textbook of Organic Medicinal and Pharmaceutical Chemistry*, 8th ed., J. B. Lippincott Co., New York, 1982, pp. 96–98.

109. K. H. Buckel, ed., *Chemistry of Pesticides*, John Wiley & Sons, Inc., New York, 1983.

110. C. Sine, ed., *Farm Chemicals Handbook*, Vol. 75, Meister Publishing Co., Willoughby, Ohio, 1991.

111. W. T. Thomson, *Agricultural Chemicals, Book 2*, Thomson Publications, Fresno, Calif., 1989.

112. T. W. Greene and P. G. M. Wuts, *Protective Groups in Organic Synthesis*, John Wiley & Sons, Inc., New York, 1991.

113. M. Bodanszky, *Principles of Peptide Synthesis*, Springer-Verlag, Berlin, 1984.

114. E. Haslam, *Tetrahedron* **30**, 2409 (1980).

115. F. J. Karol in G. A. Somorjai, ed., *Studies with High Activity Catalyst for Olefin Polymerization*, Vol. 26, no. 3 and 4, Marcel Dekker, Inc., New York, 1984, pp. 557–559.

General References

S. Patai, ed., *The Chemistry of Carboxylic Acids and Esters*, Wiley-Interscience Inc., New York, 1969.

S. Patai, ed., *The Chemistry of Acid Derivatives*, Suppl. B, Parts 1 and 2, John Wiley & Sons, Inc., New York, 1979.

R. W. Johnson and E. Fritz eds., *Fatty Acids in Industry*, Marcel Dekker, Inc., New York, 1989.

I. O. Sutherland in D. Barton and W. D. Ollis, eds., *Comprehensive Organic Chemistry*, Vol. 2, Pergamon Press, Inc., Elmsford, N.Y., 1979, pp. 869–956.

B. M. Trost and I. Fleming, eds., *Comprehensive Organic Synthesis: Selectivity, Strategy and Efficiency in Modern Organic Chemistry*, Vol. 1–9, Pergamon Press, Inc., Elmsford, N.Y., 1991.

J. March, *Advanced Organic Chemistry*, 4th ed., John Wiley & Sons, Inc., New York, 1992.

Kwoliang D. Tau

Varadaraj Elango

Joseph A. McDonough

Hoechst Celanese Corporation

ESTROGENS.

See HORMONES.

ESTRONE.

See HORMONES.

ETHANE.

See HYDROCARBONS.

ETHANOIC ACID.

See ACETIC ACID AND DERIVATIVES, ACETIC ACID.

ETHANOL

Ethanol [64-17-5] or ethyl alcohol, CH₃CH₂OH, is one of the most versatile oxygen-containing organic chemicals because of its unique combination of properties as a solvent, a germicide, a beverage, an antifreeze, a fuel, a depressant, and especially as a chemical intermediate for other organic chemicals. The use of fermentation-derived ethanol as an automotive fuel additive to enhance octane and reduce emissions has seen explosive growth over the last 12 years so that it now accounts for over 60% of the 4.5 × 10⁹-L ethanol market in the United States.

The name alcohol is a generic name derived from two Arabic words, *al* and *kohl*, that described a finely ground antimony powder used by Oriental women to darken their eyebrows. The name evolved to indicate a great degree of fineness or purity and gradually became specific for ethyl alcohol, "spirits of wine rectified to the highest degree" (1). Ethyl alcohol is, of course, well known as a constituent of alcoholic beverages (see BEVERAGE SPIRITS, DISTILLED).

As a beverage ethanol had been prepared and used long ago by the Egyptian pharaohs (2,3). Some indication of the antiquity of the knowledge of ethyl alcohol is the fact that Noah is believed to have built a vineyard in which he grew grapes that he fermented into a type of alcoholic beverage (4).

Physical Properties

Ethyl alcohol under ordinary conditions is a volatile, flammable, clear, colorless liquid. Its odor is pleasant, familiar, and characteristic, as is its taste when suitably diluted with water. The most amazing property of ethanol is the volume shrinkage that occurs when it is mixed with water, or the volume expansion that occurs when it is mixed with gasoline. One volume of ethanol plus one volume of water results in only 1.92 volumes of mixture.

The physical and chemical properties of ethyl alcohol are primarily dependent upon the hydroxyl group. This group imparts polarity to the molecule and gives rise to intermolecular hydrogen bonding. These two properties account for the differences between the physical behavior of lower molecular weight alcohols and that of hydrocarbons of equivalent weight. Infrared spectrographic studies (5) have shown that, in the liquid state, hydrogen bonds are formed by the attraction of the hydroxyl hydrogen of one molecule and the hydroxyl oxygen of a second molecule. This bonding makes liquid alcohol behave as though it were largely dimerized. This behavior is analogous to that of water, which, however, is more strongly bonded and appears to exist in liquid clusters of more than two molecules. The association of ethyl alcohol is confined to the liquid state; in the vapor state it is monomeric.

A summary of physical properties of ethyl alcohol is presented in Table 1. Detailed information on the vapor pressure, density, and viscosity of ethanol can be obtained from References 6–14. A listing of selected binary and ternary azeotropes of ethanol is compiled in Reference 15.

Table 1. Physical Properties of Ethanol^a

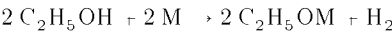
Property	Value
freezing point, °C	−114.1
normal boiling point, °C	78.32
critical temperature, °C	243.1
critical pressure, kPa ^b	6383.48
critical volume, L/mol	0.167
critical compressibility factor, z_c in $PV = znRT$	0.248
density, d_4^{20} , g/mL	0.7893
refractive index, n_D^{20}	1.36143
$\Delta n_D/\Delta t$, 20–30°C, per °C	0.000404
surface tension, at 25°C, mN/m (dyn/cm)	23.1
viscosity, at 20°C, mPa·s (cP)	1.17
solubility in water, at 20°C	miscible
heat of vaporization, at normal boiling point, J/g ^c	839.31
heat of combustion, at 25°C, J/g ^c	29676.69
heat of fusion, J/g ^c	104.6

flammable limits in air, vol %	
lower	4.3
upper	19.0
autoignition temperature, °C	423.0
flash point, closed-cup, °C	14
specific heat, at 20°C, J (g·°C) ^c	2.42
thermal conductivity, at 20°C, W (m·K)	0.170
dipole moment, liq at 25°C, C·m ^d	5.67 × 10 ^{−30}
magnetic susceptibility at 20°C	0.734 × 10 ^{−6}
dielectric constant at 20°C	25.7

^a Refs. 6–14.
^b To convert kPa to atm, divide by 101.3.
^c To convert J to cal, divide by 4.184.
^d To convert C·m to debye, divide by 3.336 × 10^{−30} (esu = D × 10^{−18}).

Chemical Properties

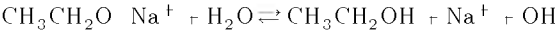
The chemistry of ethyl alcohol is largely that of the hydroxyl group, namely, reactions of dehydration, dehydrogenation, oxidation, and esterification. The hydrogen atom of the hydroxyl group can be replaced by an active metal, such as sodium, potassium, and calcium, to form a metal ethoxide (ethylate) with the evolution of hydrogen gas (see ALKOXIDES, METAL).



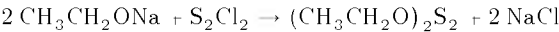
Sodium ethoxide [141-52-6] can be prepared by the reaction of absolute ethyl alcohol and sodium, or by refluxing absolute ethyl alcohol with anhydrous sodium hydroxide (16):



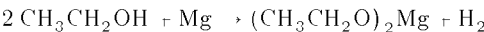
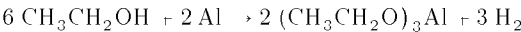
Commercially, water is removed by azeotropic distillation with benzene (17). Sodium ethoxide precipitates upon addition of anhydrous acetone (18). This strong base hydrolyzes readily to give ethyl alcohol and sodium and hydroxyl ions.



Sodium ethoxide can also be prepared by the reaction of sodium amalgam with ethyl alcohol.
Sodium ethoxide is used in organic synthesis as a condensing and reducing agent. The reaction between sodium ethoxide and sulfur monochloride yields diethyl thiosulfite (19).

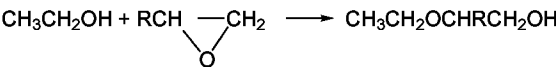


Barbiturates (Veronal, Barbital, Luminal, Amytal), ethyl orthoformate, and other chemicals are produced commercially from sodium ethoxide.
Aluminum and magnesium also react to form ethoxides, but the reaction must be catalyzed by amalgamating the metal (adding a small amount of mercury).



Well-cleaned aluminum filings react at room temperature in the presence of mercuric chloride (20,21). In an autoclave, metallic aluminum and ethyl alcohol react without a catalyst at 120°C (22). The reaction can also be promoted by the addition of sodium ethoxide (23). Aluminum should be avoided as a material of construction for ethanol service.

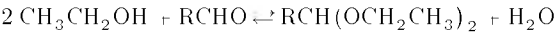
Other reactions involving the hydrogen atom of the hydroxyl group in ethyl alcohol include the opening of epoxide rings to form hydroxy ethers,



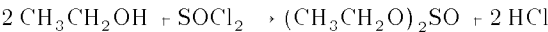
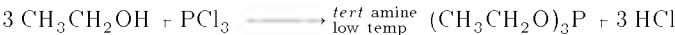
and the addition to acetylene [74-86-2] to form ethyl vinyl ether [104-92-2].



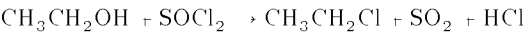
These reactions are carried out in the presence of acidic and basic catalysts. The acid-catalyzed addition of ethyl alcohol to acetylene or to a vinyl ether produces acetals (diethers of 1,1-dihydroxyethane). The acid-catalyzed reaction of ethyl alcohol with an aldehyde or ketone also gives acetals.



The hydroxyl group can be replaced by halogens from inorganic acid halides or phosphorus halides to give two different products, ethyl esters of the acid and ethyl halide (1). Phosphorus trihalides and thionyl chloride, SOCl₂, are used to make triethyl phosphite, diethyl sulfite, or ethyl chloride. The ethyl chloride yield is reduced by formation of mixed alkyl esters of phosphites, such as ethyl or diethyl phosphite, CH₃CH₂OP(OH)₂ and (CH₃CH₂O)₂POH, respectively. *Triethyl phosphite or diethyl sulfite reaction*



Ethyl chloride reaction

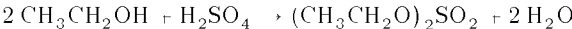
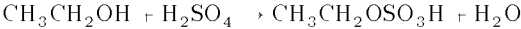


Reaction of ethanol with phosphorus trichloride produces mainly triethyl phosphite; ethyl bromide is the principal product of reaction with phosphorus tribromide. However, reaction conditions strongly affect the composition of reaction products.
The halogen acids also produce alkyl halides.

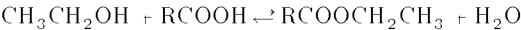


The halogen influences the rate of reaction, and, in general, the order of reactivity is HI > HBr > HCl. Important uses of ethyl chloride include the manufacture of tetraethyllead and ethylcellulose. Ethyl bromide can be used to produce ethyl Grignard reagent and various ethyl amines.

Esterification. Esters are formed by the reaction of ethanol with inorganic and organic acids, acid anhydrides, and acid halides. If the inorganic acid is oxygenated, eg, sulfuric acid, nitric acid, the ester has a carbon–oxygen linkage that is easily hydrolyzed (24–26).

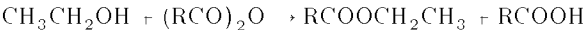


Organic esters are formed by the elimination of water between an alcohol and an organic acid (see [ESTERIFICATION](#)).

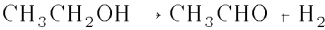


The reaction is reversible and reaches equilibrium slowly. Generally, acidic catalysts are used, such as strong sulfuric acid, hydrochloric acid, boron trifluoride, and *p*-toluenesulfonic acid (27). Batchwise and continuous processes are used for the esterification reaction.

Ethyl alcohol also reacts with acid anhydrides or acid halides to give the corresponding esters.

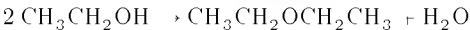
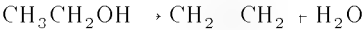


The direct conversion of ethyl alcohol to ethyl acetate is believed to take place via acetaldehyde and its condensation to ethyl acetate (Tishchenko reaction) (28–34).



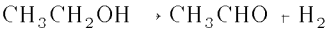
An ethyl acetate yield of ~24% is obtained using a copper oxide catalyst with 0.1–0.2% thorium at 350°C.

Dehydration. Ethyl alcohol can be dehydrated to form ethylene or ethyl ether.

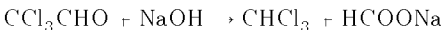
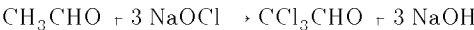
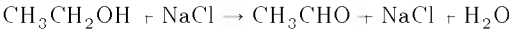


Generally, both ethylene and ethyl ether are formed to some extent, but the conditions can be altered to favor one reaction or the other.

Dehydrogenation. The dehydrogenation of ethyl alcohol to acetaldehyde can be effected by a vapor-phase reaction over various catalysts.



Haloform Reaction. Ethyl alcohol reacts with sodium hypochlorite to give chloroform [67-66-3] (haloform reaction).



Similarly, bromoform [75-25-2], CHBr₃, and iodoform [75-47-8], CHI₃, are obtained from sodium hypobromite and hypoiodite, respectively. Ethyl alcohol is the only primary alcohol that undergoes this reaction.

Concentration Effects. The reactivity of ethyl alcohol–water mixtures has been correlated with three distinct alcohol concentration ranges (35,36). For example, the chromium trioxide oxidation of ethyl alcohol (37), the catalytic decomposition of hydrogen peroxide (38), and the sensitivities of colloidal particles to coagulation (39) are characteristic for ethyl alcohol concentrations of 25–30%, 40–60%, and above 60% alcohol, respectively. The effect of various catalysts also differs for different alcohol concentrations (35).

Manufacture

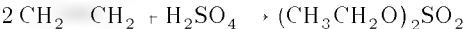
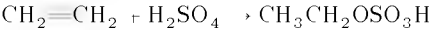
Industrial ethyl alcohol can be produced synthetically from ethylene [74-85-1], as a by-product of certain industrial operations, or by the fermentation of sugar, starch, or cellulose. The synthetic route supplies most of the industrial market in the United States. The first synthesis of ethanol from ethylene occurred in 1828 in Michael Faraday's lab in Cambridge (40).

There are two main processes for the synthesis of ethyl alcohol from ethylene. The earliest to be developed (in 1930 by Union Carbide Corp.) was the indirect hydration process, variously called the strong sulfuric acid–ethylene process, the ethyl sulfate process, the esterification–hydrolysis process, or the sulfation–hydrolysis process. This process is still in use in Russia. The other synthesis process, designed to eliminate the use of sulfuric acid and which, since the early 1970s, has completely supplanted the old sulfuric acid process in the United States, is the direct hydration process. This process, the catalytic vapor-phase hydration of ethylene, is now practiced by only three U.S. companies: Union Carbide Corp. (UCC), Quantum Chemical Corp., and Eastman Chemical Co. (a Division of Eastman Kodak Co.). UCC imports crude industrial ethanol, CIE, from SADAF (the joint venture of SABIC and Pecten [Shell]) in Saudi Arabia, and refines it to industrial grade.

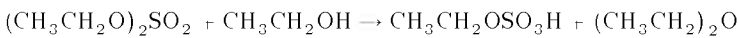
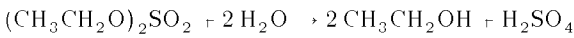
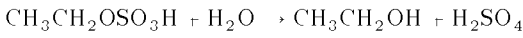
Other synthetic methods have been investigated but have not become commercial. These include, for example, the hydration of ethylene in the presence of dilute acids (weak sulfuric acid process); the conversion of acetylene to acetaldehyde, followed by hydrogenation of the aldehyde to ethyl alcohol; and the Fischer-Tropsch hydrocarbon synthesis. Synthetic fuels research has resulted in a whole new look at processes to make lower molecular weight alcohols from synthesis gas.

Indirect Hydration (Esterification–Hydrolysis) Process. The preparation of ethanol from ethylene by the use of sulfuric acid is a three-step process (Fig. 1):

- (1) Absorption of ethylene in concentrated sulfuric acid to form monoethyl sulfate (ethyl hydrogen sulfate) and diethyl sulfate:



- (2) Hydrolysis of ethyl sulfates to ethanol:



(3) Reconcentration of the dilute sulfuric acid.

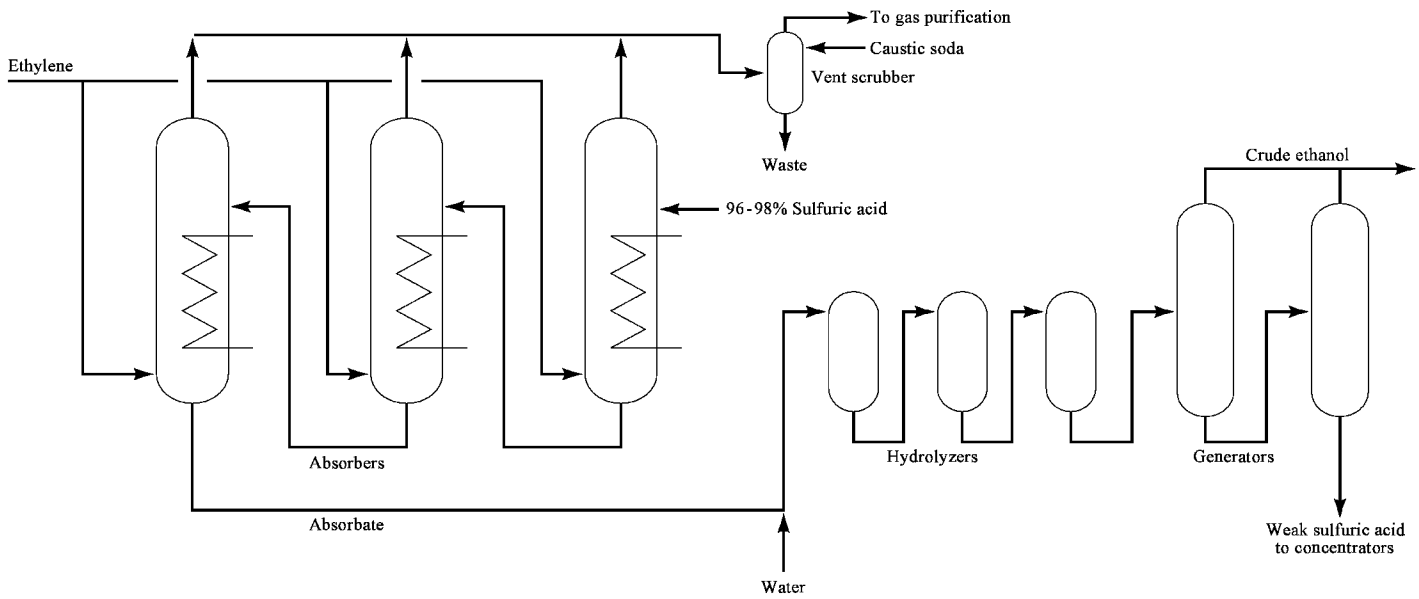


Fig. 1. Indirect hydration process for ethanol manufacture.

The hydrocarbon feedstock contains 35–95% ethylene; the remaining gases are methane and ethane. Certain unsaturated hydrocarbons are undesirable as their presence leads to the formation of secondary alcohols.

The absorption is carried out by countercurrent passage of ethylene through 95–98% sulfuric acid in a column reactor at 80°C and 1.3–1.5 MPa (180–200 psig) (41). The absorption is exothermic, and cooling is required (42) to keep the temperatures down and thereby limit corrosion problems. The absorption rate increases when ethyl hydrogen sulfate is present in the acid (43–46). This increase is attributed to the greater solubility of ethylene in ethyl hydrogen sulfate than in sulfuric acid.

The effects of various catalysts (47,48), contaminants (49,50), acid concentration (51), temperature (52), and pressure (53–57) on the rate of absorption have been studied. The patent literature indicates that absorption can be improved by making the contact between the gaseous ethylene and liquid sulfuric acid more efficient (58–61), by suitable design of the absorption tower (62), and by various combinations of absorption and hydrolysis (63–68).

The absorbate containing the mixed ethyl sulfates is hydrolyzed with enough water to give an approximately 50–60% aqueous sulfuric acid solution. The hydrolysis mixture is separated in a stripping column to give dilute sulfuric acid bottoms and a gaseous alcohol–ether–water mixture overhead. The overhead mixture is washed with water or dilute sodium hydroxide and then purified by distillation (63,65,66,68,69).

Diethyl ether is the principal by-product of the reaction of ethyl alcohol with diethyl sulfate. Various methods have been proposed to diminish its formation (70–72), including separation of diethyl sulfate from the reaction product. Diethyl sulfate not only causes an increase in ether formation but is also more difficult to hydrolyze to alcohol than is ethyl hydrogen sulfate. The equilibrium constant for the hydrolysis of ethyl hydrogen sulfate is independent of temperature, and the reaction rate is proportional to the hydrogen ion concentration (73–75).

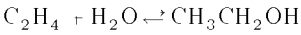
The reconcentration of dilute (50–60%) sulfuric acid is one of the more costly operations in the manufacture of ethanol by this process. An acid reboiler, followed by a two-stage vacuum evaporation system, raises acid concentration to about 90%. The 90% acid is then brought to 96–98% strength by fortification with 103% oleum (fuming sulfuric acid).

The buildup of carbonaceous materials in the sulfuric acid presents one of the most serious problems of acid concentration (76–80). Acid concentration also presents a corrosion problem. The vessels are mild steel lined with lead or brick; the steam heating elements are composed of silicon, iron, or tantalum, and pipelines are generally constructed of lead (81).

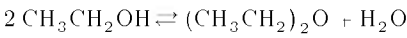
Direct Hydration of Ethylene. Hydration of ethylene to ethanol via a liquid-phase process catalyzed by dilute sulfuric acid was first demonstrated more than a hundred years ago (82). In 1923, the passage of an ethylene–steam mixture over alumina at 300°C was found to give a small yield of acetaldehyde, and it was inferred that this was produced via ethanol (83). Since the late 1920s, several industrial concerns have expressed interest in producing ethanol synthetically from ethylene over solid catalysts. However, not until 1947 was the first commercial plant for the manufacture of ethanol by catalytic hydration started in the United States by Shell; the same process was commercialized in the United Kingdom in 1951.

There are two main process categories for the direct hydration of ethylene to ethanol. Vapor-phase processes contact a solid or liquid catalyst with gaseous reactants. Mixed-phase processes contact a solid or liquid catalyst with liquid and gaseous reactants. Generally, ethanol is produced by a vapor-phase process; mixed-phase processes are used for the analogous hydration of propylene to 2-propanol. Important exceptions to these two generalizations exist, but the discussion that follows emphasizes technology associated with the commercially important vapor-phase direct hydration of ethylene.

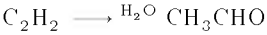
Chemistry. The stoichiometric equations pertinent to the vapor-phase hydration of ethylene over a catalyst support impregnated with phosphoric acid have been summarized (84).



Diethyl ether can be formed from the alcohol, or conversely, ether can be hydrated to ethanol.



Acetaldehyde, a deleterious by-product, is most likely formed from trace amounts of acetylene in the feed (85,86).



The acetaldehyde is particularly undesirable as it leads to the formation of crotonaldehyde (87), an impurity that adversely affects ethanol quality even at parts per million levels.

$$\Delta F_f = 28.6 T - 9.740$$

where f = fugacity, K_f = equilibrium constant based on fugacity of components, and F_f = free energy based on fugacity.

On the basis of a calculated equilibrium constant (125), the equilibrium conversions have been calculated at various temperatures (Fig. 2), pressures, and water-ethylene ratios (125).

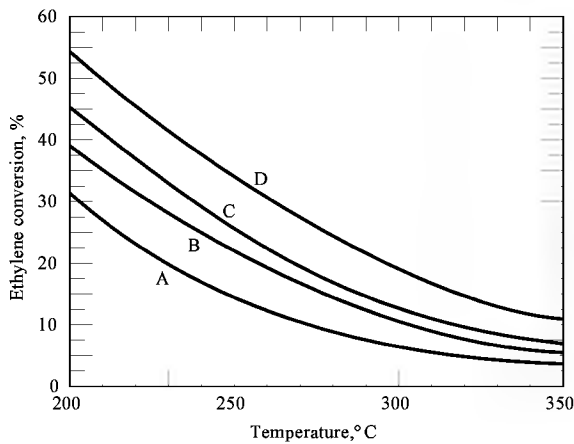


Fig. 2. Calculated conversions at equilibrium in vapor-phase hydration with equimolar ethylene and water at A, 5.1 MPa; B, 8.1 MPa; C, 10.1 MPa; and D, 15.2 MPa (125). To convert MPa to atm, divide by 0.101.

Effect of Process Variables. Several investigators (103,106) have studied the effects of the following reaction parameters:

- (1) An increase in pressure causes a corresponding increase in ethanol production rate. Higher pressures also increase polymer formation; hence, there is little practical advantage to be gained above a certain limit.
- (2) An optimum temperature exists at which the ethanol production rate is maximal. Ethylene conversion is limited by catalyst activity at lower temperatures and by equilibrium considerations at higher temperatures.
- (3) An optimum ethylene-to-water ratio exists that gives a maximum ethanol production rate. However, as expected, the highest ethylene conversion is obtained at the lowest ethylene-to-water mole ratio.
- (4) An increases in space velocity increases the ethanol production rate, but at the expense of incurring higher recycling costs.

Hibernia-Chemie has described a vapor-phase process that passes fresh and recycled 85 wt % phosphoric acid over a catalyst of hydrochloric acid-leached bentonite impregnated with phosphoric acid. Catalyst activity was claimed to remain constant over a period of one year at the following conditions (126):

temperature, °C	265
pressure, MPa (atm)	7.115 (70.23)
space velocity, h ⁻¹	1727
mole ratio, ethylene–water at inlet	1.2
conversion per pass, %	6.18
yield per pass, %	5.98
selectivity, %	96.8

Process. Figure 3 shows a simplified flow diagram for the process previously employed by Union Carbide to produce ethanol by the direct hydration of ethylene (127). An ethylene-rich gas is combined with process water, heated to the desired reaction temperature, and passed through a fixed-bed catalytic reactor to form ethanol. The vapor leaving the reactor is slightly hotter than the feed because the reaction is exothermic. The reactor product is cooled by heat exchange with the reactor feed stream and is separated into liquid and vapor streams. The liquid stream goes to the ethanol refining system, and the vapor stream is scrubbed with water to remove ethanol. The washed gas, mostly unreacted ethylene, is enriched with fresh ethylene feed and recycled to the reactor. A small vent or purge stream is removed from the recycled ethylene to prevent buildup of unwanted impurities in the gas cycle.

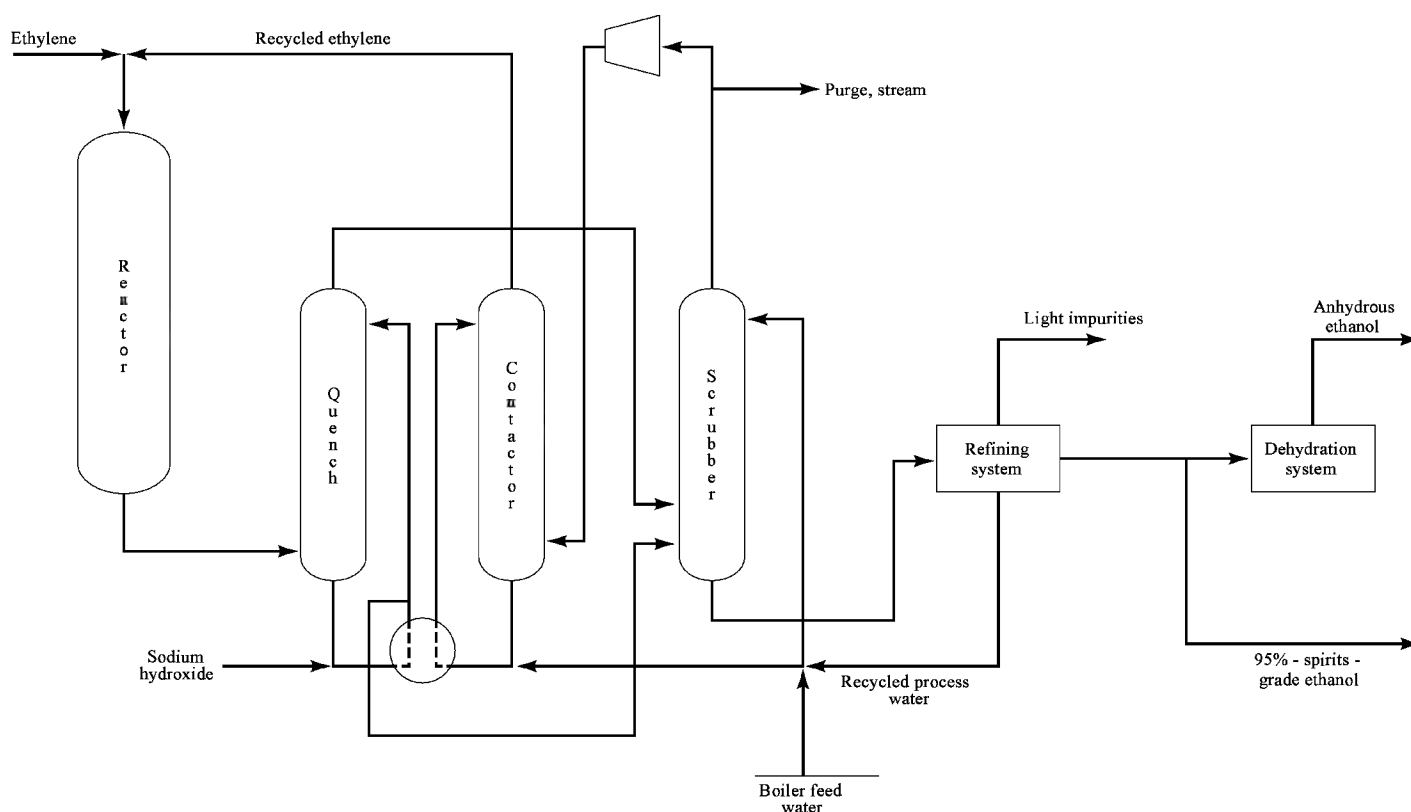


Fig. 3. Direct hydration process for ethanol manufacture.

The liquid product streams are fed to a distillation system to remove the light impurities and to recover the ethanol as a 95° volume ethanol–water azeotrope. To produce anhydrous ethanol, the ethanol–water azeotrope is fed to a dehydration system.

An advantage provided by this process is the recycling of process water recovered in the refining stills to the reaction system. This reduces the amount of boiler feed water to less than one-fifth of the total amount of water fed to the reactor. The investment and operating cost required for supporting boiler-feed water facilities are correspondingly reduced. Recycling process water also reduces the amount of water discharged to the sewer, thus decreasing the ethanol losses and the load on pollution-abatement facilities.

Several patents (128–130) deal with methods for preventing the formation of deposits in heat exchangers, reducing corrosion and avoiding the need for corrosion-resistant materials. Copper is widely used for lining the reactors and for piping, and some heat exchangers are made of phosphor bronze. Eastman Kodak Co. (131) advocates the use of a stainless-steel-clad reactor lined with overlapping copper curtains or shingles for corrosion resistance. A Hibernia-Chemie patent (96) claims that a copper lining in the reactor has a limited life and can promote the formation of cuprene. A porous carbon brick lining with a Cu–Ag alloy between the brick and the reactor wall is described (126).

Other Methods of Preparation. In addition to the direct hydration process, the sulfuric acid process, and fermentation routes to manufacture ethanol, several other processes have been suggested. These include the hydration of ethylene by dilute acids, the hydrolysis of ethyl esters other than sulfates, the hydrogenation of acetaldehyde, and the use of synthesis gas. None of these methods has been successfully implemented on a commercial scale, but the route from synthesis gas has received a great deal of attention since the 1974 oil embargo.

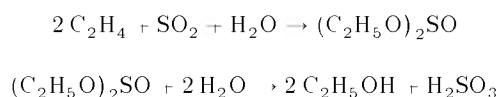
Hydration of Ethylene Using Dilute Acids. A review of the early work on the hydration of ethylene using dilute acids, the weak acid process, has been given (48). The reaction is favored by the use of low temperatures and high pressures. Temperatures in the range of 150–250°C are the most frequently quoted (132–137), although temperatures as low as 80°C have been reported (138).

Equilibrium constants for the hydration of ethylene were measured (139) using a 3 mol % sulfuric acid solution at 176–307°C and 8.1–26.3 MPa (80–260 atm). Above 250°C and 20.3 MPa (200 atm) polymerization of ethylene became important, and below 220°C ethyl ether formation was significant. Both high temperature and high acid concentration favor the reduction of sulfuric acid to sulfur dioxide (140); 65° acid is reduced by ethylene at above 180°C. To control the severe corrosion problems encountered, a tantalum-plated reactor has been recommended (141). Loss of acid catalyst in the off-gas can be compensated by adding ethyl sulfate in the feed (142). Operation at pressures from atmospheric to 2.5 MPa (25 atm) using a battery of reactors have been described (143). Ethylene was absorbed in the acid solution during countercurrent passage through the battery at 130°C.

Hydration of Ethyl Ether. Using the same type of acid catalysts as in the hydration of ethylene to ethanol, ethyl ether can be hydrated to the alcohol. Catalysts that have been used for the hydration of ether include phosphoric acid (144), sulfuric acid (145,146), hydrochloric acid (147), metallic oxides (141,148,149) and silicates (150). Sulfuric acid concentrations ranging from 5–25° at 200°C (144) to 63–70° at 110–135°C and 1.01–1.42 MPa (10–14 atm) (148) have been claimed.

Aluminum oxide has been the most widely used catalyst (151). At 320°C and 1.01–1.42 MPa, 50–66° conversion to alcohol based on the ether was obtained. Ethanol produced by the direct hydration of ether generally has a foul odor owing to the presence of polymeric hydrocarbon material, which can be removed by washing the aqueous alcohol with ether (152).

Hydrolysis of Ethyl Esters. The hydrolysis of esters (other than ethyl sulfates) is not a commercial route for producing ethanol. An indirect hydration of ethylene actually takes place during the proposed (153) hydrolysis of ethyl sulfite catalyzed by silver sulfate.



The hydrolysis of ethyl acetate, prepared by the reaction of ethylene with acetic acid under pressure (154), and the hydrolysis of the ethyl ester of chlorosulfonic acid (155) have been considered and found to be of little industrial importance.

Hydrogenation of Acetaldehyde. Acetaldehyde made from acetylene can be hydrogenated to ethanol with the aid of a supported nickel catalyst at 150°C (156). A large excess of hydrogen containing 0.3° of oxygen is recommended to reduce the formation of ethyl ether. Anhydrous ethanol has also been made by hydrogenating acetaldehyde over a copper-on-pumice catalyst (157).

Oxidation of Hydrocarbons. Ethanol is one of a variety of oxygen-containing compounds produced by the oxidation of hydrocarbons. Ethanol is reported to be obtained in a yield of 51° by the slow combustion of ethane (158,159). When propane is oxidized at 350°C under a pressure of 17.2 MPa (170 atm) (160,161), 8° of the oxygen is converted to ethanol. Lower conversions to ethanol are obtained by oxidizing butane. Other oxidation systems used to produce ethanol and acetaldehyde (162–164) and methods for separating the products have been described in the patent literature.

Synthesis Gas. Since petroleum prices rose abruptly in 1974, the production of ethanol from synthesis gas, a mixture of carbon monoxide and hydrogen, has received considerable attention. The use of synthesis gas as a base raw material has the same drawback as fermentation technology: low yields limited by stoichiometry.

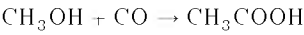


Its appeal lies in the fact that synthesis gas can be produced from trash, municipal sewage, scrap wood, sawdust, newsprint, or other waste. The early work of Fischer and Tropsch on methanol synthesis showed that ethanol could be obtained in the process (165) and that by certain modifications the proportion of ethanol in the product could be increased (166). The literature concerning this method is extensive (167–176). The conditions that favor ethanol formation are 125–175°C and 1.42 MPa (14 atm) in the presence of reduction catalysts such as powdered iron.

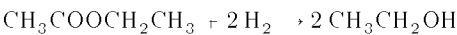
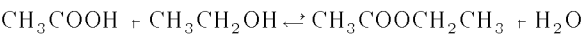
Ethanol can also be obtained by the reaction of methanol with synthesis gas at 185°C and under pressure (6.9–20.7 MPa or 68–204 atm) in the presence of a cobalt octacarbonyl catalyst (177). However, although ethanol was the primary product, methyl formate, methyl, propyl and butyl acetates, propyl and butyl alcohols, and methane were all present in the product.

A Belgian patent (178) claims improved ethanol selectivity of over 62%, starting with methanol and synthesis gas and using a cobalt catalyst with a halide promoter and a tertiary phosphine. At 195°C, and initial carbon monoxide pressure of 7.1 MPa (70 atm) and hydrogen pressure of 7.1 MPa, methanol conversions of 30% were indicated, but the selectivity for acetic acid and methyl acetate, useful by-products from this reaction, was only 7%. Ruthenium and osmium catalysts (179,180) have also been employed for this reaction. The addition of a bicyclic trialkyl phosphine is claimed to increase methanol conversion from 24% to 89% (181).

Rather than converting methanol directly to ethanol, two processes have been announced that go through the intermediate step of converting the methanol to acetic acid by rhodium-catalyzed carbonylation.



One process (182) esterifies the acetic acid with ethanol (or methanol) and then converts the ester to alcohol by hydrogenolysis in the vapor phase over a copper–zinc catalyst.



The other process (183) hydrogenates the acetic acid directly to ethanol.

References 184 to 188 provide general reviews of the fuel alcohol from synthesis gas research.

Fermentation Ethanol. Fermentation (qv), one of the oldest chemical processes known to man, is used to make a variety of products, including fuel, foods, flavorings, beverages, pharmaceuticals, and chemicals. At present, however, many of the simpler products such as ethanol, are synthesized from petroleum feedstocks at lower costs. Ethanol production by fermentation, excluding that for beverages, had been declining in the United States since synthetic ethanol was introduced in the 1930s, because of the low cost and assured availability of ethylene. The quadrupling of the selling price of crude petroleum by the Organization of Petroleum Exporting Countries (OPEC) in 1973 had a profound impact on fermentation processes for producing ethanol. The U.S. Department of Energy set objectives to develop methods to derive fuels economically from sugar crops and corn, to evaluate the potential feasibility of the various methods, and to suggest means of practical application. This program resulted in many economic studies and laboratory research programs. Interest then waned as the price of oil dropped, until 1979 when the Islamic revolution in Iran caused another oil crisis.

State and federal tax subsidies and loan guarantees fueled the growth of fermentation ethanol capacity in the early 1980s. In 1980, loan guarantees of nearly \$342 × 10⁶ from the Farmers Home Administration were approved for 15 new plants in 14 states to produce 931 × 10⁶ L per year for fuel as part of the synfuels program (189). In the mid-1980s the phase-out of lead as an octane enhancer in gasoline kept the fuel ethanol program alive. The 1990 Clean Air Act requirements for oxygenates and renewal of the Federal tax rebates worth \$0.16 per liter of ethanol pumped new life into the fuel ethanol program in the late 1980s. The estimated 1991 demand for fuel ethanol was over 3.0 × 10⁹ L (190).

Most of the fermentation ethanol is made from corn even though before World War II molasses was the chief feedstock. Ethanol can be made from a variety of agricultural products such as other grains, sugar cane and beets, fruit, whey, and sulfite waste liquor. Studies are underway to ferment garbage to ethanol. Generally, most of the agricultural products mentioned command higher prices as foods, and others, eg, potatoes, are uneconomical because of their low ethanol yield and high transportation cost. The use of fermentation ethanol in the industrial market depends on the availability and cost of the carbohydrate relative to the availability and cost of ethylene, and the economics and government policies relating to the fuel ethanol market. Current U.S. producers of fermented industrial alcohol include ADM Corp., Georgia Pacific, Grain Processing Corp., Midwest Grain Products, Inc., and Union Carbide Corp.

In some European countries (eg, Belgium, France, Italy, and the Netherlands) fermentation is still important in producing industrial ethanol. In others (eg, England and Germany) synthetic ethanol predominates (191). In Japan the synthetic capacity has been shut down except for refining imported crude, while the fermentation capacity continues to operate. Fuel ethanol programs in other countries, such as Brazil, have spurred the growth of worldwide fermentation ethanol capacity.

Ethanol can be derived by fermentation processes from any material that contains sugar or compounds that can be converted to sugar. The many and varied raw materials used in the manufacture of ethanol via fermentation are conveniently classified under three types of agricultural raw materials: sugar, starches, and cellulose materials. Sugars (from sugar cane, sugar beets, molasses, or fruit) can be converted to ethanol directly. Starches (from grains, potatoes, or root crops) must first be hydrolyzed to fermentable sugars by the action of enzymes from malt or molds. Cellulose (from wood, agricultural residues, or waste sulfite liquor from pulp and paper mills) must likewise be converted to sugars, generally by the action of mineral acids. Once simple sugars are formed, enzymes from yeast can readily ferment them to ethanol (192). Because fermentation ethanol has been thoroughly and repeatedly discussed in the literature, the coverage here is illustrative rather than comprehensive, with special emphasis on the potential raw materials for ethanol production of the future (see also BEVERAGE SPIRITS, DISTILLED).

Sugars. Prior to the late 1970s, the most widely used sugar for ethanol fermentation was blackstrap molasses (192–203) which contains about 35–40 wt % sucrose, 15–20 wt % invert sugars such as glucose and fructose, and 28–35 wt % of nonsugar solids. Blackstrap (derived from Java and the Dutch word *stroop*, meaning syrup) is collected as a by-product of cane sugar manufacture. The molasses is diluted to a mash containing ~10–20 wt % sugar. After the pH of the mash is adjusted to about 4–5 with mineral acid, it is inoculated with the yeast, and the fermentation is carried out nonaseptically at 20–32°C for about 1–3 d. The fermented beer, which typically contains ~6–10 wt % ethanol, is then sent to the product recovery and purification section of the plant.

The direct fermentation of sugar cane juice, sugar beet juice, beet molasses (a by-product in the production of beet sugar), fresh and dried fruits, cane sorghum, whey, and skim milk had been considered as a means of obtaining ethanol, but none of these raw materials could compete economically with molasses. Although the manufacture of ethanol from the sugar-containing waste products of the fruit industry appears to be a highly desirable operation, particularly as a means of reducing stream pollution in the vicinity of canning plants, such production is costly because of the need to remove most of the water (as much as 97%) contained in the waste product.

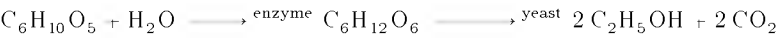
The results of the sugar-crop research on agronomics and fuels conversion undertaken by Batelle's Columbus Laboratories (194) lists several merits of sugar cane as a candidate energy resource. Sugar cane, a renewable raw material, is renowned for its agricultural productivity, and its juice is directly fermentable to ethanol. On the other hand, sugar cane products are valuable in food and feed applications and their conversion to chemicals and energy can be considered an underutilization of their potential value. By the year 2000 the energy needs of the United States are expected to exceed ~10¹⁷ kJ (100 quads or 10¹⁷ Btu) (195,196) and sugar cane is expected to contribute less than 2 × 110¹⁵ kJ (2 quads) of energy.

In 1975, Brazil embarked upon the ambitious ProAlcohol program for fermentation ethanol manufacture (198,199) from sugar cane to reduce the

country's dependence on foreign oil, and to modernize and make more competitive Brazil's sugar mills in a lagging international sugar market. Gasohols have been used in Brazil since the early 1930s. The number of alcohol distilleries grew to over 500 while the number of big sugar plantations grew to over 1000 (204). This growth was fueled by government subsidies and the production of automobiles that ran totally on ethanol. As the price of oil dropped in the 1980s, the production of fuel ethanol and the expansion of the sugar cane industry stagnated. Production of fuel ethanol was 11.9×10^9 L in 1989, representing a demand shortfall of 1.4×10^9 L. This shortfall was caused by lack of sugar cane, as the estimated distilling capacity is 15×10^9 L. The \$18 $\times 10^9$, 15-year program appeared headed for a phase-out in 1990, until Iraq's invasion of Kuwait caused renewed fears of oil supply shortfalls and soaring oil prices. After Operation Desert Storm restored peace in the Middle East, oil prices have returned to preinvasion levels, and it remains to be seen in what direction the ProAlcohol program is headed.

The subject of fermentation alcohol has always been of considerable interest to several tropical countries, but until the oil crisis of 1973, other than Brazil (197), only India appeared to appreciate the importance of fermentation alcohol as a strategic material in its economy. Ethanol prices in India have been maintained at an extremely low level by processing cane molasses, which has been a waste product of negligible value (197).

Starches. In the United States, all potable alcohol, most fermentation industrial alcohol, and most fuel alcohol is currently made principally from grains; corn is the principal feedstock for fuel alcohol. Fermentation of starch from grain is somewhat more complex than fermentation of sugars because starch must first be converted to sugar and then to ethanol. This process was known to the ancient Egyptians and Mesopotamians who brewed beer almost 5000 years ago (202). The simplified equations for the conversion of starch to ethanol are



Starch is converted enzymatically to glucose either by diastase present in sprouting grain or by fungal amylase. The resulting dextrose is fermented to ethanol with the aid of yeast, producing CO₂ as a coproduct. Other by-products depend on the type of process.

The basic process steps for converting corn into ethanol are degermination, milling, separation of starch-bearing endosperm from hulls, slurrying–liquefaction, hydrolysis of starch to sugar, fermentation, distillation, and dehydration (205). The hydrolysis or saccharification is usually carried out with an amylase enzyme and the fermentation is usually by the yeast *Saccharomyces cerevisiae*. Corn can be prepared for fermentation by three different processes: whole-grain grinding, dry milling, and wet milling. Whole-grain grinding requires the least capital investment but generates only one by-product, distillers-dried grains (DDG) which is used to supplement animal feed. This process does not include the degermination step. A dry mill has the next lowest capital investment, but generates only the lower valued by-products including DDG. A wet mill is often built in conjunction with a plant that produces high fructose corn syrup from part of the starch, which is used as a beverage sweetener. Other by-products from a wet mill are corn gluten and corn gluten meal, which are used for animal feeds. While more expensive to build and operate, the greater variety of by-products can help offset the price of corn. Most of the fuel alcohol plants use the dry-mill or wet-mill process.

The fuel alcohol program has spawned a tremendous amount of research aimed at improving the cost and efficiency of the corn process. Three promising technologies for lowering operating costs are (1) substituting yeast with a high temperature bacteria such as *Zymomonas mobilis*; (2) using a permeable membrane to separate dissolved solids and some of the water before distillation; and (3) immobilizing the yeasts and enzymes in the wet-mill process to provide continuous processes with higher productivities (206). Commercialization of these technologies might save \$0.013 L in operating costs. Some researchers estimate even higher savings, up to \$0.066 L, for immobilized enzyme–yeast processes (207). Most fermentation processes are operated batchwise, but some continuous processes have already been commercialized (208).

Cellulosic Materials. Over 900 $\times 10^6$ metric tons of carbohydrate-containing cellulosic wastes are generated annually. The technology for converting this material into ethanol is available, but the stoichiometry of the process is disadvantageous. Even if each step in the process of the conversion of cellulose to ethanol proceeded with 100% yields, almost two-thirds of the mass would disappear during the sequence, most of it as carbon dioxide in the fermentation of glucose to ethanol. This amount of carbon dioxide leads to a disposal problem rather than to a raw material credit (209).

Starch (qv) and cellulose (qv) are both polymers of glucose, but cellulose is much more difficult to hydrolyze to the sugar. Its structure is more crystalline which protects the internal bonds from hydrolysis, and cellulose in plants is protected by lignin (qv), a polyphenolic material that forms a seal around the cellulose for further protection against hydrolysis (210). Cellulosic wastes also contain substantial amounts of hemicellulose (qv), which is a polymer of pentoses. The aqueous mineral acids used to hydrolyze the cellulose to glucose destroy much of the sugars, particularly the pentoses, in the process. Nevertheless, a 1978 study claimed that forests could theoretically provide 50% of the oil and gas used by U.S. utilities, replacing 20% of annual fossil fuel consumption (211). None of this has taken place, but research has continued. A new process utilizing low temperature hydrolysis to separate cellulose from paper has been licensed, with plans to construct a plant in Germany (212). The process uses electrodialysis rather than diffusion dialysis to recover hydrochloric acid for reuse. Other new ways of reducing the cost of converting cellulosic wastes from wood, newspapers, and municipal garbage into glucose include the use of less corrosive acids and reduced hydrolysis time. One way of making cellulose wastes more susceptible to hydrolysis is by subjecting them to a short burst of high energy electron beam radiation (213). Hydropulping of cellulose feedstocks followed by a 10 μ s burst from a 3×10^6 -eV electron-beam accelerator is claimed to reduce the time of hydrolysis by dilute acid from hours to seconds.

An alternative to acid hydrolysis is the use of enzymes. Although they avoid the corrosion problems and loss of fuel product associated with acid hydrolysis, enzymes have their own drawbacks (see ENZYME APPLICATIONS, INDUSTRIAL). Enzymatic hydrolysis slows as the glucose product accumulates in a reaction vessel. This end-product inhibition eventually halts the hydrolysis unless some way is found to draw off the glucose as it is formed. In mid-1978, Gulf Oil researchers described the simultaneous enzymatic hydrolysis of cellulose and fermentation of the resulting glucose to ethanol, removing glucose as it is formed and overcoming the problem of product inhibition of hydrolysis (211). Mutated strains of the common soil mold *Trichoderma viride* can process 15 times as much cellulose as natural strains. The results have been encouraging; in some cases, cellulose from sawdust, bark, and effluent streams from the pulp and paper industries have produced ethanol in yields approaching 100% of the theoretical value. A sequential hydrolysis process has been proposed (210), in which the hemicellulose is hydrolyzed to pentose by aqueous sulfuric acid and then separated for fermentation. The remaining ligno-cellulose is then pretreated with the solvent cadoxen (5–7% cadmium oxide in 28% aqueous ethylenediamine) which breaks the lignin seal to allow enzyme-catalyzed hydrolysis to glucose. Pentose fermentation to ethanol is more difficult and this process allows the two sugars to be fermented separately. Research is underway on genetically engineered bacteria for cellulose conversion (214). A novel pretreatment for municipal solid waste consists of soaking the waste in high pressure liquid ammonia (215). An instant pressure release opens up the fiber structure so that enzymes can more easily penetrate and digest the cellulose. A 25% increase in the amount of waste digested to sugar is claimed. Steam explosion has been reported as an effective pretreatment for the enzymatic hydrolysis of wood (qv) and agricultural residues (216).

Recovery and Purification

Various distillation and equipment modifications are used to ensure a pure water azeotrope of ethanol (95% by weight ethanol). A review of the patent literature concerning the many methods used for the purification of ethanol obtained from various synthetic processes is given in Reference 217. Some phosphoric acid is vaporized, or entrained as liquid, from the reactor. During recovery the reactor effluent is cooled, causing the condensation of acid on exchanger and piping walls, creating a potential corrosion problem. Corrosion can be prevented by spraying the effluent with caustic solutions so that the resultant condensate has a pH in the range 6.5–7.5. A second benefit is that this treatment aids in achieving an acceptable odor in the purified product (218). Eastman (219) and Hibernia (128) describe the use of a spray directly into the reactor effluent; Shell (218) partially cools the effluent in a heat exchanger before contacting with a spray.

The specifications for an industrial alcohol depend on the intended use. Probably for this reason noncomparable purification schemes are frequently found in the patent literature (220–224). Premium-grade ethanol has low water content, low odor ratings, and high permanganate-time-test values. Odor is improved by the partial hydrogenation of ethanol using Raney nickel catalyst (225) or by contact with unglazed porcelain and iron (226). The permanganate test is extremely sensitive to the presence of aldehydes. One ppm of crotonaldehyde or 2 ppm of sorbaldehyde is reported to decrease the test time from 60 to 30 min (218). This sensitivity suggests that purification by distillation alone is inadequate for achieving high permanganate test times (218).

Carbonyl-containing and unsaturated materials are removed by treatment with sodium borohydride (227,228) and boric acid (229). Other methods used to remove carbonyl impurities include treatment with hydroxyl amine hydrochloride, potassium permanganate, or *N*-hydroxybenzenesulfonamide (229).

Purification schemes have generally emphasized the following techniques: (1) extractive distillation using water reflux to distill a large share of the impurities and concentrate the crude alcohol–water mixture; (2) efficient fractionation to produce approximately 190 proof alcohol; (3) hydrogenation to convert aldehyde impurities to alcohols, together with the use of chemicals such as inorganic bases and sodium sulfite (222); and (4) ion-exchange resins or azeotropic distillation to dehydrate 190-proof to 200-proof or absolute alcohol.

Fuel-grade ethanol does not meet the high purity requirements of the industrial market and requires further purification for crossover to industrial markets. Much research has been carried out to lower the costs of separating ethanol from fermentation processes. Fuel use requires anhydrous ethanol, less than 0.5 wt % water, so the ethanol–water azeotrope must be overcome. Conventional distillation schemes consume 50–80% of the energy used in a typical fermentation plant (230). Techniques involving extraction (231,232), critical fluid extraction (233), extractive distillation (230,234), membrane separation (235), and water adsorption by molecular sieves (236) or cellulose, carboxymethylcellulose, cornmeal, cracked corn, corn cobs, wheat straw, bagasse, starch, hemicellulose, wood chips, other grains, and other agricultural residues (237) have been reported.

Manufacturing Costs

The cost of producing industrial ethanol depends on the location of the manufacturing plant; the design, type, and degree of modernization of equipment; the kind of raw material used; the price paid for the raw material; the relative labor costs represented; the scale of production; and the total investment. On a raw material basis, each kilogram of ethanol requires 0.6 kg ethylene, 2 kg sugar, 3.3 kg corn, or 4 kg molasses. If all other cost factors were equivalent, the economics of fermentation ethanol derived from corn would compare favorably with synthetic ethanol if corn prices were 5.5 times less than ethylene on a weight basis. With ethylene at \$0.44 /kg, as it was in 1991, corn would have to be about \$0.08 /kg (about \$2 /bushel). Corn by-products that are produced by some fermentation plants offset the price of corn to a varying degree depending on their market value. It is expected that ethylene will continue to be the primary raw material for industrial ethanol, because during the early to mid-1990s most of the fermentation ethanol will go to fuel due to the impact of the Clean Air Act, and ethylene prices are not expected to significantly increase because of additional ethylene capacity coming onstream (238) (see ETHYLENE).

Inasmuch as all fermentation alcohol plants differ in some respects, no accurate, generally applicable, basic investment cost can be formulated, and no standard production technique exists. Large facilities usually produce ethanol at lower cost than small facilities because of economies of scale in both production technology and distribution of the product (205). In general, plants utilizing molasses apply a much simpler process and require less equipment than do grain plants, since use of the latter entails grain handling and by-product feed recovery operations and has higher steam requirements. Most fermentation processes are operated batchwise and much research is underway to develop more cost-effective continuous processes.

The efficient production of ethanol from sugar cane is based on using the fibrous by-products (bagasse) to generate the steam needed to crush and mill the agricultural inputs, with enough steam left over to conduct the ethanol distillation. In 1976, Batelle (194) estimated the manufacturing cost for ethanol on a 265 × 10⁶ L/yr scale. They estimated that net raw material costs (cane juice cost minus by-products value) were 48% of the total manufacturing costs, with operating costs other than raw materials equal to 36% (including energy costs of 23%) and capital charges equal to 16%.

The cost of producing ethanol from corn depends on the highly variable corn price and the selling price for the process-dependent by-products such as high protein feeds. The calculated total cost of production has ranged from a low of \$0.20 /L with the unusually high by-product prices in 1987 to \$0.37 to 0.40 /L during 1981, 1983, and 1984 (206). The price of corn has ranged from less than \$0.043 /L (\$1.50 /bushel) in 1987 to over \$0.085 /L (\$3 /bushel) in 1981, 1983, and 1984. By-product sales have offset as little as 30% of the corn cost for dry mills to as much as 90% of the corn cost for wet mills in 1987.

Operating costs for large plants other than corn costs ranged from \$0.11 /L of ethanol to 0.16. Energy was the single largest component of these costs, accounting for an average of 36%. For a plant with \$0.12 /L operating costs, the breakdown would be: energy \$0.045 /L; ingredients, personnel, and maintenance \$0.063 /L; and management, administration, insurance, and taxes \$0.016 /L. Under the latest tax laws, capital charges range from \$0.05 to 0.13 /L.

Crossover of fermentation ethanol from the fuel market to the industrial market depends mainly on the relative profit that fermenters can make in the fuel market versus the industrial market. Profit in the fuel market is highly dependent on the price of crude oil and the federal and state subsidies. With corn at \$0.085 /L (\$2 /bushel) and existing federal subsidies, ethanol from average existing fermentation plants is competitive if crude oil trades at about \$0.15 /L (\$24 /barrel) (206). With technological improvements that save \$0.013 /L in operating costs, ethanol is competitive with \$0.13 /L (\$20 /barrel) crude oil. Without the federal subsidy, ethanol would not be competitive until crude oil sold for \$0.25 /L (\$39 /barrel). Another factor affecting the profitability is the relative value of ethanol versus MTBE (methyl *tert*-butyl ether) and other oxygenates required by the Clean Air Act. The federal subsidy is still necessary to provide the profitability for this use at current oil prices (see ETHERS).

Shipment

Commercial ethyl alcohol is shipped in railroad tank cars, tank trucks, 208-L (55-gal) and 19-L (5-gal) drums, and in smaller glass or metal containers having capacities of 0.473 L (one pint), 0.946 L (one quart), 3.785 L (one U.S. gal), or 4.545 L (one Imperial gal). The 208-L drums may be of the unlined iron type. If a guarantee of more meticulous quality is desired, the drums may be lined with phenolic resin. All containers, of course, must comply with the specifications of the U.S. Department of Transportation. Both 190 proof and 200 proof ethyl alcohol are considered red label (flammable) materials by the DOT, as both have flash points below 37.8°C by the Tag closed-cup method.

Economic Aspects

Worldwide ethanol demand in 1991 was about 19 × 10⁹ L, of which the industrial demand was about 5 × 10⁹ L. The majority of the worldwide demand was for fuel use with Brazil consuming 11.9 × 10⁹ L in 1989 (239) and the United States consuming 3.6 × 10⁹ L in 1990 (240). In 1991, worldwide synthetic ethanol capacity amounted to 2.04 × 10⁹ L, with the United States having a capacity of 0.802 × 10⁹ L, 39% of the world total (Table 2). In 1991, the total ethanol capacity in the United States was 4.67 × 10⁹ L with an industrial demand of 0.908 × 10⁹ L. Fermentation alcohol and imports supplied the remainder of the industrial market not supplied by synthetic alcohol.

Table 2. Worldwide Synthetic Ethanol Capacity in 1991^a

Geographic location	Capacity, L × 10 ⁶
United States	802
Union Carbide ^b	454
Quantum	250
Eastman	98
United Kingdom	374
France	133
Germany	227
Eastern Europe ^c	295
Japan	115
Korea	90

Total

2036

^a Data from various sources.

^b Union Carbide refines crude industrial alcohol imported from Saudi Arabia.

^c Includes Russia.

Foreign synthetic ethanol capacity increased dramatically in the 1970s. In 1973, U.S. producers commanded a hefty 74% of the world synthetic ethanol capacity; in 1977 the United States had only 54% of the worldwide capacity. By 1991, with the shutdown of several U.S. facilities, the United States share had dropped to 39%.

The industrial ethanol market in the United States experienced high growth rates after the introduction of synthetic ethanol in the 1930s (Fig. 4). World War II caused a rapid increase in usage for butadiene manufacture, mostly supplied by fermentation from idle beverage plants. After the war, synthetic ethanol captured virtually all of the market as it grew to a peak demand of 1.2×10^9 L in the late 1960s. Demand dropped sharply in the early 1970s when new technologies for the production of acetaldehyde, acetic acid, butanol, and ethylhexanol were commercialized. With the shutdown of several synthetic producers in the early 1980s, and the rapid buildup of fermentation ethanol capacity for the fuel market, fermentation ethanol gained an increasing share of the market through the 1980s. It is predicted that fermentation ethanol's share of the industrial market will decrease through the 1990s as more product goes into the fuel market. Growth in the industrial market was 1.2% yr in the 1980s and is projected to be 1.5–2% yr through 1995 (241). Growth in the fuel market was 27% yr over the same period, and is projected to grow at 5–10% through 1995, even though the actual growth will depend heavily on how the oxygenate requirements of the 1990 Clean Air Act are met. Some industry observers are predicting that fuel ethanol demand might double by 1995 (242). Although the addition of ethanol to gasoline reduces carbon monoxide emissions, it increases fuel volatility and thus VOCs or volatile organic compounds. These VOCs form ozone in the air in the presence of sunlight. The Clean Air Act requires states in areas where concentrations of ozone exceed the standard to control emission sources. Gasoline blends containing ethanol currently enjoy an exemption on vapor pressure (243). To overcome the VOC issue, ethanol could be converted to ETBE, ethyl *tert*-butyl ether, a proposed fuel oxygenate similar to MTBE. Usage would then depend on ethanol price (with the federal subsidy) and availability compared to methanol. Some MTBE producers are undertaking preliminary work to allow ETBE production in their MTBE plants (244).

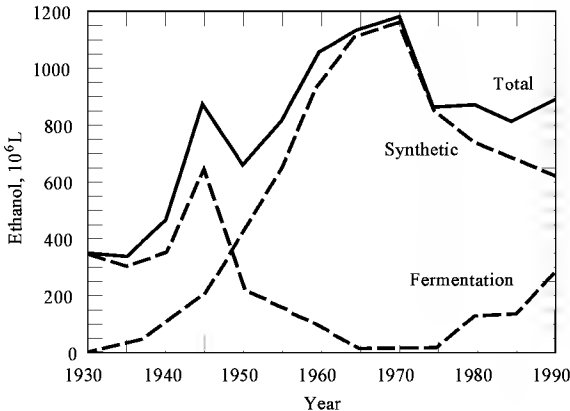


Fig. 4. U.S. Industrial ethanol market by source.

The market price of industrial ethanol has ranged from a high of \$0.47 /L for 190-proof (tax-free) to a low of \$0.06 /L over the time period of 1940–1990 (Table 3). Until the early 1960s, the price of industrial alcohol in the United States was closely related to the cost of molasses during normal world conditions and to the cost of grain during emergency times. When synthetic ethanol became the dominant source of U.S. production, greater price stability was obtained (the price of synthetic ethanol was related to the price of ethylene). The March 1992 price ranged from \$0.58 /L to \$0.61 /L (245). Market price for fuel-grade ethanol was \$0.40 /L and federal rebates to blenders reduced the net price by \$0.16 /L. Fuel-grade prices have traditionally been based on the wholesale price of gasoline plus the federal subsidy, but new formulas that recognize its value as a key oxygenate and octane component are emerging (246). Thus the price of industrial ethanol could be affected by all of these factors in the future.

Table 3. Ethanol 190- and 200-Proof Price History, \$ /L^a

Year	190-Proof		Tax per proof L, \$	200-Proof	
	Tax-free	Tax-paid		Tax-free	Tax-paid
1940	0.06	1.56	0.79	0.08	1.67
1945	0.13	4.65	2.37	0.15	4.90
1950	0.10	4.62	2.37	0.12	4.87
1955	0.10	5.37	2.77	0.12	5.67
1960	0.13	5.41	2.77	0.16	5.70
1965	0.13	5.41	2.77	0.16	5.70
1970	0.14	5.41	2.77	0.16	5.70
1975	0.26	5.54	2.77	0.28	5.83
1980	0.47	6.74	3.30	0.52	7.12
1985	0.42	6.70	3.30	0.46	7.07
1990	0.43	6.71	3.30	0.47	7.08

^a Data from various sources.

U.S. ethanol exports have been largely dependent on the world price of sugar in recent years. When sugar prices have risen to the point that it was more profitable for the Brazilian sugar mills to refine sugar cane into sugar rather than fermenting it into ethanol, Brazil would import large quantities of ethanol to meet the fuel demand (247).

U.S. ethanol imports were greatly influenced by the rapid increase of world fermentation capacity in the late 1970s and the rapid increase in U.S. demand for fuel ethanol. Imported alcohol from Brazil reached 140×10^6 L in 1980, mostly for the fuel market. In 1981, the U.S. Government passed an import duty on fuel ethanol. The duty was raised to \$0.158 /L in 1985 (248), after Brazilian imports reached 314×10^6 L in 1984. Total ethanol imports in 1984 were 568×10^6 L, with 95×10^6 L going to the industrial market, where the import duties do not apply. Another factor influencing ethanol imports

for the fuel market is the Caribbean Basin Initiative, CBI, which grants duty-free status to some fuel ethanol imports produced in the Caribbean from imported feedstocks (249). The initiative was established in 1983 as the Caribbean Basin Economic Recovery Act to provide Caribbean and Central American nations with economic security. Interpretation of this act has led to many disputes in the U.S. Congress (250–254). Fuel ethanol imports are important to the industrial ethanol market, since one of the factors affecting cross-over from the fuel market is the supply demand balance in the United States.

Units, Specifications, and Test Methods

Units. The alcohol content of spirits is usually given in terms of proof, an archaic term inherited from early distillers of fermentation alcohol. In England, the proof was to pour some of the spirit over gunpowder, and ignite the spirit; at or above a limiting concentration (11 parts of alcohol by volume to 10 parts of water), the gunpowder would explode. Because volumes were at the time much easier to measure accurately than weights, this measurement of alcohol persisted, even though there is a considerable volume change on mixing ethyl alcohol with water.

In the United States the proof is twice the alcohol content by volume, thus 190-proof alcohol contains 95° º ethyl alcohol by volume (Table 4). According to Federal statute, "... proof spirits shall be held to be that alcoholic liquor which contains one-half of its volume of alcohol of a specific gravity of 0.7939 at 15.5°C." A gallon (3.785 L) of proof spirits can be made by mixing 0.5000 gal (1.8925 L) of absolute alcohol with 0.5373 gal (2.0337 L) water; it contains 42.49° º alcohol by weight. A wine gallon is a measure of quantity, 231 in.³ (3.785 L), of any proof. A proof gallon (tax gallon) is a wine gallon of 100-proof spirits, or its equivalent.

Table 4. Conversion of U.S. Proof to Alcohol by Volume, Alcohol by Weight, and British Proof^a

U.S. proof ^b	Alcohol, vol ° º ^b	Alcohol wt ° º	British proof
0	0.0	0.00	100.0
2	1.0	0.80	98.4
4	2.0	1.59	96.8
6	3.0	2.39	95.2
8	4.0	3.19	93.6
10	5.0	4.00	91.9
20	10.0	8.05	83.5
30	15.0	12.14	75.0
40	20.0	16.27	66.1
50	25.0	20.44	57.0
60	30.0	24.67	48.0
70	35.0	28.97	39.3
80	40.0	33.36	30.6
90	45.0	37.86	21.7
100	50.0	42.49	12.9
110	55.0	47.24	4.0
120	60.0	52.15	4.8 ^c
130	65.0	57.21	13.5 ^c
140	70.0	62.44	22.3 ^c
150	75.0	67.87	31.3 ^c
160	80.0	73.53	39.9 ^c
170	85.0	79.44	48.6 ^c
180	90.0	85.69	57.3 ^c
190	95.0	92.42	66.0 ^c
200	100.0	100.0	76.0 ^c

^a Underproof, unless otherwise indicated (proof is given a value of zero).

^b At 15.5°C.

^c Overproof.

The determination of the proof (the alcohol content) is usually made by measuring specific gravity with hydrometers at the standard temperature of 15.5°C, since only very small amounts of impurities other than water are usually present.

Specifications. The specifications for ethyl alcohol are designed with sufficient latitude to allow for the two principal means of production, synthesis from ethylene and fermentation. The requirements given by the *U.S. Pharmacopeia* (USP) and the American Chemical Society (ACS) generally form the foundation for the most widely used specifications (255,256). A tabulation of the specifications from the major alcohol producers, with consideration for the USP and ACS requirements, is shown in Table 5.

Table 5. Typical Ethanol Specifications

Requirement	190-Proof	200-Proof
specific gravity, 20–20°C, max	0.8126	0.7912
purity, vol ° º, min	95	99.9
acidity, wt ° º as CH ₃ COOH, max	0.002	0.002
nonvolatile matter, g/100 mL, max	0.0025	0.0025
miscibility with water	complete	complete
permanganate time test, minutes, min	40	25
odor	no foreign or residual	no foreign or residual
color, APHA, ^a max	10	10
water, wt ° º, max		0.1

^a American Public Health Association.

The primary producers of ethyl alcohol also market the specially denatured and completely denatured alcohols, as well as various proprietary solvents in which ethyl alcohol is the basic ingredient. These various products can also be described by rigid and descriptive specifications, but the

requirements must make allowance for the chemical and physical character of the denaturants.

The use of ethyl alcohol in some medicinal and cosmetic products requires a very meticulous grade, particularly with reference to odor. In some instances, the odor can be correlated with the concentration of certain minor impurities; in most instances it cannot be directly associated with any measurable contaminant, and the quality can be ascertained only by odor comparison with previously accepted material.

Test Methods. The most generally used means of ascertaining the purity of commercial ethyl alcohol is by specific gravity determination, sometimes referred to as alcoholometry. For this reason, the specific gravity should be determined very accurately to the fourth decimal place by means of a calibrated pycnometer or hydrometer (257). The value obtained is referred to standard U.S. Dept. of the Treasury tables relating specific gravity to alcohol content. Table 6 is extracted from more detailed government tables. Of course, this procedure is valid only for undenatured alcohol, since the purity nomograph is based on the two-component alcohol-water mixture.

Table 6. Respective Volumes of Alcohol and Water and Specific Gravity ^a

U.S. proof	Alcohol, vol	Water, vol	Apparent sp gr ^b
101	50.60	53.24	0.93320
110	55.00	48.76	0.92409
120	60.00	43.71	0.91333
130	65.00	38.60	0.90190
140	70.00	33.43	0.88986
150	75.00	28.19	0.87714
160	80.00	22.87	0.86364
170	85.00	17.46	0.84927
180	90.00	11.93	0.83362
190	95.00	6.18	0.81582
198	99.00	1.29	0.79866
200	100.00	0.00	0.79365

^a Table 6 may also be used to determine the quantity of water needed to reduce the strength of ethanol by a definite amount. Divide the alcohol in the given strength by the alcohol in the required strength, multiply the quotient by the water in the required strength, and subtract the water in the given strength from the product. The remainder is the number of liters (gallons) of water to be added to 378.5 L (100 gal) of spirits of the given strength to produce a spirit of a required strength.

^b 15.6 15.6°C.

The other analytical methods necessary to control the typical specification given in Table 5 are, for the most part, common quality-control procedures. When a chemical analysis for purity is desired, acetylation or phthalation procedures are commonly employed. In these cases, the alcohol reacts with a measured volume of either acetic or phthalic anhydride in pyridine solution. The loss in titratable acidity in the anhydride solution is a direct measure of the hydroxyl groups reacting in the sample. These procedures are generally free from interference by other functional groups, but both are affected adversely by the presence of excessive water, as this depletes the anhydride reagent strength to a level below that necessary to ensure complete reaction with the alcohol. Both procedures can be adapted to a semimicro- or even microscale determination.

Of course, acetylation and phthalation are not selective for ethyl alcohol, but determine any reactive OH group present. A survey of methods applicable to the determination of ethyl alcohol has been published (258). A method using a solution of acetyl chloride in toluene has been described (259). Owing to both the reactivity and volatility of the reagent, special techniques and careful handling are required for accurate analyses, and in general the method is less satisfactory than that using acetic anhydride. To overcome the difficulties of using acetyl chloride solutions, 3,5-dinitrobenzoyl chloride can be used, and after hydrolysis, titration with tetrabutylammonium hydroxide (260) is carried out. Pyromellitic dianhydride has also been used as an acylating agent (261) in the presence of aldehydes.

Colorimetric methods have been successfully used for determining trace amounts of ethanol. Ammonium hexanitratocerate(IV) has been used as a reagent (262) and for continuous automatic analysis. Alcohols form colored complexes with 8-hydroxyquinoline and vanadic compounds. The absorbance of these complexes, measured at 390 μm has been used to provide an analytical procedure (263).

Acetone (264) and ethyl acetate (265) have been used as internal standards for the gas chromatographic determination of ethanol. This technique and enzymatic methods (266) have become widely used for estimation of ethanol in blood samples. Capillary gas chromatographs are replacing packed column chromatographs because of their higher sensitivity and better reproducibility. Automatic versions of the alcohol dehydrogenase method have been devised (267) and evaluated (268). The content of alcohol in wines has been measured using a differential refractometer (269). An automatic online head-space gas chromatographic method has been developed for quantitative determination of fermentation ethanol (270). The determination of ethanol in exhaled air by such equipment as the Breathalyzer (271) and Alcotest (272) depends on color changes in a solid adsorbent.

Health and Safety Considerations

Ethyl alcohol is a flammable liquid requiring a red label by the DOT and Coast Guard shipping classifications; its flash point is 14°C (Tag, closed cup). Vapor concentrations between 3.3 and 19.0% by volume in air are explosive. Liquid ethyl alcohol can react vigorously with oxidizing materials. Ethyl alcohol has found wide application in industry, and experience shows that it is not a serious industrial poison (273–275). If proper ventilation of the work environment is maintained, there is little likelihood that inhalation of the vapor will be hazardous.

The threshold limit value for ethyl alcohol vapor in air has been set at 1000 ppm for an 8-h time-weighted exposure by the ACGIH (1989 listing). The minimum identifiable odor of ethyl alcohol has been reported as 350 ppm. Exposure to concentrations of 5,000–10,000 ppm result in irritation of the eyes and mucous membranes of the upper respiratory tract and, if continued for an hour or more, may result in stupor or drowsiness. Concentrations of this latter order of magnitude have an intense odor and are almost intolerable to begin with, but most people can become acclimated to the exposure after a short time. Table 7 gives the effects of exposure to even heavier concentrations.

Table 7. Ethanol Vapor Concentration and Its Effects in Humans ^a

Concentration		Effects in humans
mg L of air	ppm vol of air	
10–20	5,300–10,640	some transient coughing and smarting of the eyes and nose which disappear after 5–10 min; not comfortable but tolerable
30	15,960	continuous lacrimation and marked coughing; could be tolerated but with discomfort
40	21,280	just tolerable for short periods
>40	>21,280	intolerable and suffocating for even short periods

^a Ref. 276.

Ethyl alcohol is oxidized completely to carbon dioxide and water in the body, thus it is not a cumulative poison. An average adult is able to oxidize each hour the equivalent of 19 g of 100 proof whiskey (277). Less than 10⁰ % of the absorbed alcohol is excreted, chiefly in urine, measurably in expired air, and detectably in sweat (278). Alcohol poisoning and alcohol intoxication are almost invariably the result of using alcohol as a beverage, rather than inhalation as a vapor. About 75–80 g of ingested alcohol will produce symptoms of intoxication in an average (70-kg) person. About 150–200 g will cause stupor, and 250–500 g may be a fatal dose. Severe hypoglycemia in certain individuals following a heavy drinking bout is now well established (279).

Because alcohol intoxication may be simulated by many pathologic conditions, including diabetic acidosis, the postconvulsive depression of epilepsy, uremia, head injuries, and poisonings by any other central nervous depressant and some stimulants (280), a diagnosis of acute alcoholism should not be made casually; chemical testing of blood, urine, or expired air is always desirable.

Some authorities question whether drunkenness can result from the inhalation of ethyl alcohol vapors. Experience has demonstrated that in any event such intoxication is indeed rare (281). There is no concrete evidence that the inhalation of ethyl alcohol vapor will cause cirrhosis. Liver function is definitely impaired during alcohol intoxication (282), making the subject more susceptible to the toxic effects of chlorinated hydrocarbons.

Repeated exposure to ethyl alcohol results in the development of a tolerance as evidenced by decreasing symptomatic reactions. It has been demonstrated that the symptoms of exposure are less clear and the time required to produce them is greater in subjects accustomed to alcohol. There is no proof, however, of physiological adaptation in humans in terms of metabolic changes or resistance to cellular injuries. The subject of the interaction of alcohol with other drugs has received much attention (277).

Uses

Industrial ethanol is one of the largest-volume organic chemicals used in industrial and consumer products. The main uses for ethanol are as an intermediate in the production of other chemicals (Table 8) and as a solvent. As a solvent, ethanol is second only to water. Ethanol is a key raw material in the manufacture of drugs, plastics, lacquers, polishes, plasticizers, perfumes, and cosmetics. Around 1960, manufacture of ethanol was the top consumer of ethylene in the United States, but since 1965 it has rated below manufacture of ethylene oxide and polyethylene.

Table 8. Consumption of Ethanol as a Chemical Intermediate ^a, 10⁶ L

Chemical	1950	1960	1970	1980	1990
acetaldehyde	330	592	306	30	
acetic acid	27	17	16		
ethyl acetate	24	27	23	30	7
ethyl acrylate	na	16	46	57	98
ethylamine	na	na	26	53	66
ethyl chloride	16	24	16	4	
glycol ethers	na	32	76	82	18
vinegar ^b	23	42	63	68	87
other	53	60	34	49	89
Total	473	810	606	373	365

^a Refs. 1, 241, 285.

^b Dilute solution of acetic acid.

Over the last 30 years, ethanol's role as a solvent has increased sharply, while its role as a chemical intermediate has declined. In 1990, 59⁰ % of the 890 × 10⁶ L demand was used for solvents and the remaining 41⁰ % was used for chemical intermediates (283). In 1960, solvents accounted for only 24⁰ % of the demand. The 1990 solvent uses were toiletries and cosmetics, 33⁰ %; coatings, inks, and proprietary blends, 29⁰ %; detergents and household cleaners, 14⁰ %; external pharmaceuticals, 7⁰ %; insecticides and disinfectants, 7⁰ %; and miscellaneous, 10⁰ %. Ethanol demand for solvent applications has been fairly stable in recent years, growing at an average annual rate of 2⁰ %. VOC regulations could impact its solvent use, particularly in areas like California, where ethanol in aerosols like hair spray and deodorants have come under scrutiny.

Ethanol has been used in the United States to a considerable extent as an antifreeze but it has largely been replaced for such use by ethylene glycol. The use of fermentation alcohol in automotive fuel has been discussed above. Several tropical countries consider ethanol, produced from regenerable resources, an attractive petrochemical feedstock.

Denatured Ethanol. For hundreds of years alcoholic beverages have been taxed all over the world to generate government revenue. When ethanol emerged as a key industrial raw material, the alcohol tax was recognized as a burden to many essential manufacturing industries. To lift this burden, the Tax-Free Industrial and Denatured Alcohol Act of 1906 was passed in the United States. The U.S. Treasury, Bureau of Alcohol, Tobacco, and Firearms (BATF), now oversees the production, procurement, and use of ethanol in the United States.

The concern of the government is to prevent tax-free industrial ethanol from finding its way into beverages. To achieve this end, the regulations call for a combination of financial and administrative controls (bonds, permits, and scrupulous record keeping) and chemical controls (denaturants that make the ethanol unpalatable). Regulations establish four distinct classifications of industrial ethanol. The classifications with the most stringent financial and administrative controls call for little or no chemical denaturants. The classifications that call for the most effective chemical denaturants require the least financial and administrative controls. For a list of denaturants currently authorized, see Reference 284.

Completely Denatured Alcohol. Completely denatured alcohol (CDA) escapes the involved financial and administrative controls required of the other classifications of industrial ethanol. No tax is applied, no bond is required, no permit is needed to enable a customer to purchase CDA. Requirements for records by both producer and user are minimal. These simplified regulations are possible because CDA is denatured with substances that render it totally unfit for beverage purposes. It is also unsuitable where odor is objectionable. CDA and products made from it are, however, governed by special labeling requirements of the BATF. Repackaging of completely denatured alcohol is permitted as long as labeling requirements are met.

Proprietary Solvents and Special Industrial Solvents. Proprietary solvents and special industrial solvents are made with specially denatured alcohols according to the formulas authorized by the BATF. They can be purchased by customers without payment of tax, without posting a bond for tax, and without securing a permit from the BATF. Suppliers are required, however, to notify the BATF of the name, address, type of business, and approximate annual requirements and intended end use for any user buying in bulk.

Proprietary solvents may be repackaged for retail, wholesale, and industrial sales. Retail sales of special industrial solvents are prohibited. Agents may repackage special industrial solvents for wholesale and industrial sales, but only in drum quantities and with the producer's label. Special labeling requirements of the BATF apply to both proprietary and special industrial solvents.

Specially Denatured Alcohol. Specially denatured alcohols (SDA) are formulations of ethanol containing denaturant substances that generally render them unfit for beverage use but do not limit their use in specified applications. To use a specially denatured alcohol, a manufacturer must apply to the BATF, giving quantitative formulas and processes. Specimen labels and a sample of the finished product are also required. Then, the prospective user must obtain a bond for the total amount of specially denatured alcohol on hand or in transit at any given time.

Pure Ethanol. Undenatured ethanol can be bought on either a tax-free or tax-paid basis. Approved educational, scientific or medical organizations and public agencies can buy tax-free alcohol. Use and withdrawal permits are required, as are a bond and detailed records. Resale of tax-free ethanol is prohibited. Approved industrial uses for pure tax-paid ethanol include pharmaceuticals, cosmetics, flavoring extracts, and foods.

Chemicals Derived From Ethanol

Ethanol's use as a chemical intermediate (Table 8) suffered considerably from its replacement in the production of acetaldehyde, butyraldehyde, acetic acid, and ethylhexanol. The switch from the ethanol route to those products has depressed demand for ethanol by more than 300×10^6 L (80×10^6 gal) since 1970. This decrease reflects newer technologies for the manufacture of acetaldehyde and acetic acid, which is the largest use for acetaldehyde, by direct routes using ethylene, butane (173), and methanol. Oxo processes (qv) such as Union Carbide's Low Pressure Oxo process for the production of butanol and ethylhexanol have totally replaced the processes based on acetaldehyde. For example, U.S. consumption of ethanol for acetaldehyde manufacture declined steadily from 50% in 1962 to 37% in 1964 and none in 1990. Butadiene was made from ethanol on a large scale during World War II, but this route is no longer competitive with butadiene derived from petroleum operations.

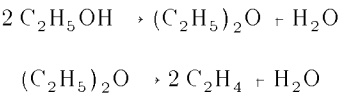
Acetaldehyde. Until the early 1970s, the main use of industrial ethanol was for the production of acetaldehyde [75-07-0]. By 1977, the ethanol route to acetaldehyde had largely been phased out in the United States as ethylene and ethane became the preferred feedstocks for acetaldehyde production (286–304). Acetaldehyde usage itself has also changed; two primary derivatives of acetaldehyde, acetic acid, and butanol, are now produced from feedstocks other than acetaldehyde. Acetaldehyde is still produced from ethanol in India.

There are two ways to produce acetaldehyde from ethanol: oxidation and dehydrogenation. Oxidation of ethanol to acetaldehyde is carried out in the vapor phase over a silver or copper catalyst (305). Conversion is slightly over 80% per pass at reaction temperatures of 450–500°C with air as an oxidant. Chloroplatinic acid selectively catalyzes the liquid-phase oxidation of ethanol to acetaldehyde giving yields exceeding 95%. The reaction takes place in the absence of free oxygen at 80°C and at atmospheric pressure (306). The kinetics of the vapor and liquid-phase oxidation of ethanol have been described in the literature (307,308).

The reaction kinetics for the dehydrogenation of ethanol are also well documented (309–312). The vapor-phase dehydrogenation of ethanol in the presence of a chromium-activated copper catalyst at 280–340°C produces acetaldehyde in a yield of 89% and a conversion of 75% per pass (313). Other catalysts used include neodymium oxide and samarium hydroxide (314).

Ethylene. Where ethylene is in short supply and fermentation ethanol is made economically feasible, such as in India and Brazil, ethylene is manufactured by the vapor-phase dehydration of ethanol. The production of ethylene [74-85-1] from ethanol using naturally renewable resources is an active and useful alternative to the pyrolysis process based on nonrenewable petroleum. This route may make ethanol a significant raw material source for producing other chemicals.

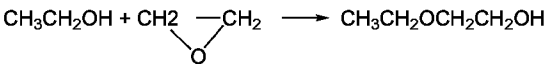
Dehydration of ethanol has been effected over a variety of catalysts, among them synthetic and naturally occurring aluminas, silica-aluminas, and activated alumina (315–322), hafnium and zirconium oxides (321), and phosphoric acid on coke (323). Operating space velocity is chosen to ensure that the two consecutive reactions,



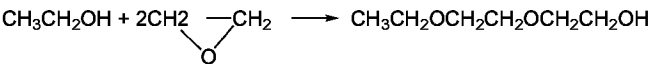
go to completion, avoiding the need to recover and recycle unreacted ethanol. The dehydration is endothermic, and temperature is a critical operating parameter; high temperatures produce aldehydes and low temperatures, ethers. The catalyst is usually regenerated with steam and air every few weeks to remove carbon deposits.

A fluidized-bed catalytic reactor system developed by C. E. Lummus (323) offers several advantages over fixed-bed systems in temperature control, heat and mass transfer, and continuity of operation. Higher catalyst activity levels and higher ethylene yields (99% compared to 94–96% with fixed-bed systems) are accomplished by continuous circulation of catalyst between reactor and regenerator for carbon burn-off and continuous replacement of catalyst through attrition.

Glycol Ethers. The addition of one mole of ethylene oxide [75-21-8] to ethanol gives ethylene glycol monoethyl ether [109-86-4].



Addition of two moles of oxide gives the monoethyl ether of diethylene glycol.



The oxide–alcohol route is the only commercially important route to glycol ethers now in use. Anhydrous alcohols must be used; otherwise the water present forms contaminating glycols.

The reactions are highly exothermic. Under liquid-phase conditions at about 200°C, the overall heat of reaction is 83.7 to 104.6 kJ mol (20 to 25 kcal mol) ethylene oxide reacting (324). The opening of the oxide ring is considered to occur by an ionic mechanism with a nucleophilic attack on one of the epoxide carbon atoms (325). Both acidic and basic catalysts accelerate the reactions, as does elevated temperature. The reaction kinetics and product distribution have been studied by a number of workers (326,327).

Ethylene oxide (qv), propylene oxide (qv), butylene oxide, and other epoxides react with ethanol to give a variety of liquid, viscous, semiwax, and solid products. These products are used in the coatings industry as solvents, and as paints, antioxidants, corrosion inhibitors, and special-purpose polymers. Recent concerns about the health effects of ethanol containing glycol ethers have led to the decline in the production of these compounds.

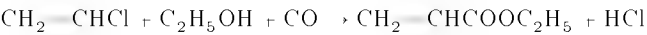
Vinegar. Dilute solutions of alcohol as fermented worts are oxidized by air at 30–40°C in the presence of various organisms such as (328) *Mycoderma aceti*, *B. aceti*, and *B. xylinus*, to produce dilute acetic acid as vinegar. Vinegar based on synthetic ethanol has a fully acceptable aroma and taste (329) and has gained a healthy share of the market for vinegar used in such products as pickles, ketchup, and mustard. However, vinegar based on fermentation ethanol is gaining back some of this market because of the "all natural ingredients" trend in advertising.

Ethylamines. Mono-, di-, and triethylamines, produced by catalytic reaction of ethanol with ammonia (330), are a significant outlet for ethanol. The vapor-phase continuous process takes place at 1.38 MPa (13.6 atm) and 150–220°C over a nickel catalyst supported on alumina, silica, or silica–alumina. In this reductive amination under a hydrogen atmosphere, the ratio of the mono-, di-, and triethylamine product can be controlled by recycling the unwanted products. Other catalysts used include phosphoric acid and derivatives, copper and iron chlorides, sulfates, and oxides in the presence of acids or alkaline salts (331). Piperidine can be ethylated with ethanol in the presence of Raney nickel catalyst at 200°C and 10.3 MPa (102 atm), to give *N*-ethylpiperidine [766-09-6] (332).

Ethyl Acrylate. The esterification of acrylic acid is a primary use for ethanol. Acrylic acid can also react with either ethylene or ethyl esters of sulfuric acid.

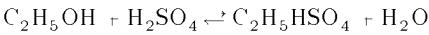


These processes have supplanted the condensation reaction of ethanol, carbon monoxide, and acetylene as the principal method of generating ethyl acrylate [140-88-5] (333). Acidic catalysts, particularly sulfuric acid (334–338), are generally effective in increasing the rates of the esterification reactions. Care is taken to avoid excessive polymerization losses of both acrylic acid and the esters, which are accentuated by the presence of strong acid catalysts. A synthesis for acrylic esters from vinyl chloride (339) has also been examined.

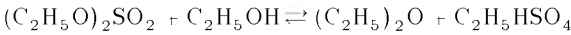


Vinyl chloride reacts at 270°C at >6.9 MPa (68 atm) with ethanol and carbon monoxide in the presence of a cobalt and palladium catalyst to give ethyl acrylate in 17% yield.

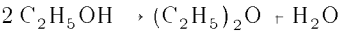
Ethyl Ether. Most ethyl ether is obtained as a by-product of ethanol synthesis via the direct hydration of ethylene. The procedure used for production of diethyl ether [60-29-7] from ethanol and sulfuric acid is essentially the same as that first described in 1809 (340). The chemical reactions involved in the production of ethyl ether by the indirect ethanol-from-ethylene process are like those for the production of ether from ethanol using sulfuric acid.



A more recent concept (341) is that ether is principally derived from the reaction between diethyl sulfate and ethanol.



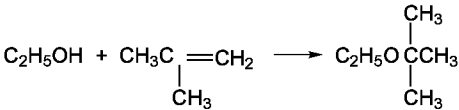
An alternative route for the formation of diethyl ether from ethanol is by dehydration:



The reaction is catalyzed by all but the weakest acids. In the dehydration of ethanol over heterogeneous catalysts, such as alumina (342–346), ether is the main product below 260°C; at higher temperatures both ether and ethylene are produced. Other catalysts used include silica–alumina (347,348), copper sulfate, tin chloride, manganous chloride, aluminum chloride, chrome alum, and chromium sulfate (349,350).

Ethyl Vinyl Ether. The addition of ethanol to acetylene gives ethyl vinyl ether [104-92-2] (351–355). The vapor-phase reaction is generally run at 1.38–2.07 MPa (13.6–20.4 atm) and temperatures of 160–180°C with alkaline catalysts such as potassium hydroxide and potassium ethoxide. High molecular weight polymers of ethyl vinyl ether are used for pressure-sensitive adhesives, viscosity-index improvers, coatings and films; lower molecular weight polymers are plasticizers and resin modifiers.

Ethyl *tert*-Butyl Ether. Ethanol can react with isobutylene to form ETBE much the same way as methanol is now processed into MTBE, methyl *tert*-butyl ether (see ETHERS).



Ethyl *t*-butyl ether [637-92-3] (ETBE) is a viable octane enhancer and oxygenate for fuel and does not increase the vapor pressure of gasoline as ethanol does when blended. Although there is no current production of ETBE, this could change now that the government has extended the tax credit for ethanol in gasolines to ETBE (356). ETBE is more desirable than ethanol to gasoline refiners and blenders because it is easier to transport and is not ruined by water. Ethanol in gasoline absorbs water when transported through pipelines and tanks to local terminals. Too much water absorption causes phase separation.

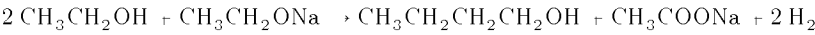
Ethyl Acetate. The esterification of ethanol by acetic acid was studied in detail over a century ago (357), and considerable literature exists on determinations of the equilibrium constant for the reaction. The usual catalyst for the production of ethyl acetate [141-78-6] is sulfuric acid, but other catalysts have been used, including cation-exchange resins (358), α -fluoronitrites (359), titanium chelates (360), and quinones and their partly reduced products.

Ethyl acetate is made industrially by both batch and continuous processes (361,362). Glacial acetic acid is commonly the starting material, and any water formed during the esterification has to be removed. Sulfuric acid may be added periodically to the reactor to replace the acid lost in side reactions. The vapor-phase esterification of ethanol has also been studied extensively (363,364), but it is not used commercially. The reaction can be catalyzed by silica gel (365,366), thoria on silica or alumina (367), zirconium dioxide (368), and by xerogels and aerogels (369). Above 300°C the dehydration of ethanol becomes appreciable. Ethyl acetate can also be produced from acetaldehyde by the Tischenko reaction (370–372) using an aluminum alkoxide catalyst and, with some difficulty, by the boron trifluoride-catalyzed direct esterification of ethylene with organic acids (373).

Ethyl Chloride. Previously a significant use for industrial ethanol was the synthesis of ethyl chloride [75-00-3] for use as an intermediate in producing tetraethyllead, an antiknock gasoline additive. Ethanol is converted to ethyl chloride by reaction with hydrochloric acid in the presence of aluminum or zinc chlorides. However, since about 1960, routes based on the direct addition of hydrochloric acid to ethylene or ethane have become more competitive (374,375).

Other Derivatives and Reactions. The vapor-phase condensation of ethanol to give acetone has been well documented in the literature (376–385); however, acetone is usually obtained as a by-product from the cumene (qv) process, by the direct oxidation of propylene, or from 2-propanol.

The Guerbet reaction (386–389) involving condensation of ethanol in the presence of sodium ethoxide, catalyzed by potassium hydroxide and boric anhydride (390,391) or alkaline phosphates (392), gives *n*-butanol [71-36-3]:



However, the oxo reaction starting from propylene and proceeding via the hydrogenation of butyraldehyde, has become the more widely employed commercial route for preparing *n*-butanol (see BUTYL ALCOHOLS; OXO PROCESS).

During World War II, production of butadiene (qv) from ethanol was of great importance. About 60% of the butadiene produced in the United States during that time was obtained by a two-step process utilizing a 3:1 mixture of ethanol and acetaldehyde at atmospheric pressure and a catalyst of tantalum oxide and silica gel at 325–350°C (393–397). Extensive catalytic studies were reported (398–401) including a fluidized process (402). However, because of later developments in the manufacture of butadiene by the dehydrogenation of butane and butenes, and by naphtha cracking, the use of ethanol as a raw material for this purpose has all but disappeared.

An extensive listing of 35 other reactions including alkylation, etherification, alcoholysis, and halogenation has been compiled (1) to show the versatility of ethanol as a reactant.

BIBLIOGRAPHY

"Alcohol, Industrial" in *ECT* 1st ed., Vol. 1, pp. 252–288, by R. S. Aries, Consulting Chemical Engineer; "Ethanol" in *ECT* 2nd ed., Vol. 8, pp. 422–470, by C. A. Pentz and G. A. Lescisin, Union Carbide Corp.; in *ECT* 3rd ed., Vol. 9, pp. 338–380, by P. D. Sherman, Jr., and P. R. Kavasmaneck, Union Carbide Corp.

1. L. F. Hatch, *Ethyl Alcohol*, Enjay Chemical Co., a Division of Humble Oil and Refining Co., New York, 1962.
2. C. M. Beamer, *Chem. Eng. Prog.* **43**(3), 92 (1947).
3. E. Huber, *Z. Spiritusind.* **50**, 164 (1927).
4. *Ethyl Alcohol Production Technique*, Noyes Development Corp., Pearl River, N.Y., 1964.
5. D. Hodzi, *Hydrogen Bonding*, Pergamon Press, London, 1957.
6. F. K. Beilstein, *Handbuch der Organischen Chemie*, ed. III, Vol. 1, Springer-Verlag, Berlin, 1958, pp. 1229–1236.
7. *Selected Values of Properties of Chemical Compounds*, Manufacturing Chemists' Association Research Project Tables, Chemical Thermodynamic Properties Center, Dept. of Chemistry, A & M College of Texas, College Station, Tex.

8. *Selected Properties of Hydrocarbons and Related Compounds*, American Petroleum Institute Project 44, Chemical Thermodynamic Properties Center, Dept. of Chemistry, A & M College of Texas, College Station, Tex.

9. J. Timmermans, *Physico-Chemical Constants of Pure Organic Compounds*, Elsevier Publishing Co., New York, 1950, pp. 306–312.

10. K. A. Kobe and R. E. Pennington, *Thermochemistry of Petrochemicals*, Reprint No. 44, University of Texas, Bureau of Engineering Research, Austin, Tex., Jan. 1949–Dec. 1951, pp. 56–59.

11. R. C. Reid and J. M. Smith, *Chem. Eng. Prog.* **47**(8), 415 (1951).

12. S. Young, *Sci. Proc. Dublin Soc.* **12**, 374 (1910).

13. A. R. Challoner and R. Powell, *Proc. R. Soc. (London)* **238A**, 90 (1956).

14. Union Carbide Corporation, Chemicals and Plastics Division, New York, 1965, unpublished physical property data.

15. L. H. Horsley, ed., *Azeotropic Data, Advances in Chemistry Series*, Vol. 226, American Chemical Society, Washington, D.C., 1973.

16. D. Williams and R. W. Bost, *J. Chem. Phys.* **4**, 251 (1936).

17. **U.S. Pat. 1,712,830** (May 14, 1929), L. P. Kyrides (to National Aniline Division, Allied Chemical & Dye Corp.).

18. **U.S. Pat. 1,978,647** (Oct. 30, 1934), E. T. Olson and R. H. Twinning (to Cleveland Cliffs Iron Co.).

19. A. Meuwsen, *Ber.* **68**, 121 (1935).

20. E. W. Zappi and E. Restelli, *An. Assoc. Quim. Argent.* **32**, 89 (1934).

21. F. Henle, *Ber.* **53**, 719 (1920).

22. **Brit. Pat. 454,480** (Oct. 1, 1936), (to Consortium für Elektrochemische Industrie GmbH).

23. **Ger. Pat. 602,376** (Sept. 17, 1934), J. Seib (to Deutsche Gold- und Silber-Scheideanstalt vorm. Roessler).

24. C. M. Suter and E. Oberg, *J. Am. Chem. Soc.* **56**, 677 (1934).

25. M. Gallagher and D. B. Keyes, *J. Am. Chem. Soc.* **56**, 2221 (1934).

26. **U.S. Pat. 2,831,882** (Apr. 22, 1958), C. P. Spaeth (to E. I. du Pont de Nemours & Co., Inc.).

27. H. D. Hinton and J. A. Nieuwland, *J. Am. Chem. Soc.* **54**, 2017 (1932).

28. P. Ya. Ivannikov and E. Ya. Gavrilova, *J. Chem. Ind. (USSR)* **12**, 1256 (1935).

29. B. N. Dolgov and co-workers, *Zh. Obshch. Khim.* **25**, 693 (1955).

30. B. N. Dolgov and co-workers, *J. Chem. Ind. (USSR)* **12**, 1066 (1935).

31. B. N. Dolgov and co-workers, *Org. Chem. Ind. (USSR)* **1**, 70 (1936).

32. N. M. Beizel and co-workers, *Org. Chem. Ind. (USSR)* **1**, 102 (1936).

33. **U.S. Pat. 2,004,350** (June 11, 1935), N. D. Scott (to E. I. du Pont de Nemours & Co., Inc.); **Brit. Pat. 424,284** (Feb. 18, 1935), N. D. Scott (to E. I. du Pont de Nemours & Co., Inc.).

34. H. Adkins and R. Conner, *J. Am. Chem. Soc.* **53**, 1091 (1931).

35. M. Bobtelsky, *J. Chem. Soc. Part IV*, 3615 (1950).

36. Ref. 1, p. 68.

37. M. Bobtelsky and R. Cohn, *Z. Anorg. Allg. Chem.* **210**, 225 (1933).

38. M. Bobtelsky and co-workers, *J. Am. Chem. Soc.* **67**, 966 (1945).

39. Br. Jirgensons, *Z. Phys. Chem.* **158A**, 56 (1931).

40. P. I. Spitz, *Petrochemicals, The Rise of an Industry*, John Wiley & Sons, Inc., New York, 1988, p. 71.

41. **Brit. Pat. 273,263** (Sept. 15, 1926) (to Compagnie de Bethune).

42. **U.S. Pat. 2,545,161** (Mar. 13, 1951), E. C. Morrell and R. F. Robey (to Standard Oil Development Co.); **Brit. Pat. 655,475** (July 25, 1951) (to Standard Oil Development Co.).

43. S. P. G. Plant and N. V. Sidgwick, *J. Soc. Chem. Ind. (London)*, **40**(2), 14T (1921).

44. F. Vallette, *Chim. Ind. (Paris)* **13**, 718 (1951).

45. **Brit. Pat. 221,512** (Feb. 19, 1925) (to Compagnie de Bethune).

46. P. Fritsche, *Chem. Ind. (London)* **20**, 266 (1897); **21**, 33 (1898).

47. C. Ellis, *Chemistry of Petroleum Derivatives*, Chemical Catalog Co., New York, 1934, p. 301.

48. C. Ellis, *Chemistry of Petroleum Derivatives*, Vol. II, Reinhold Publishing Corp., New York, 1937.

49. W. S. E. Hickson and K. C. Bailey, *Sci. Proc. R. Dublin Soc.* **20**, 267 (1932).

50. K. C. Bailey and W. E. Calcutt, *Sci. Proc. R. Dublin Soc.* **21**, 309 (1936).

51. B. T. Brooks, *Ind. Eng. Chem.* **27**, 278 (1935).

52. B. Neuman, *Gas Wasser* **67**, 1, 14, 53 (1924).

53. **Brit. Pat. 308,859** (Apr. 4, 1929), R. E. Slade (to Imperial Chemical Industries Ltd.).

54. **U.S. Pat. 1,885,585** (Nov. 1, 1932), B. T. Brooks (to Petroleum Chem. Corp.).

55. **U.S. Pat. 1,919,618** (June 25, 1933), B. T. Brooks (to Standard Alcohol Co.).

56. B. G. Simek, *Chem. Prunysl* **7**, 122 (1957); (through) *Erdoel Kohle* **10**, 882 (1957).

57. F. Strahler and F. Hachtel, *Brennst. Chem.* **15**, 166 (1934).

58. C. F. Tidman, *J. Soc. Chem. Ind. (London)* **40**, 103R (1921).

59. A. Damiens, *Compt. Rend.* **175**, 585 (1922).

60. A. Damiens, *Bull. Soc. Chim.* **33**, 71 (1923).

61. H. Hennell, *Ann. Chim.* **35**(2), 154 (1827).

62. **U.S. Pat. 2,755,297** (July 17, 1956), B. I. Smith and W. H. Rader (to Esso Research and Engineering Co.).

63. **U.S. Pat. 2,765,347** (Oct. 2, 1956), B. I. Smith (to Esso Research and Engineering Co.).

64. **U.S. Pat. 2,779,803** (Jan. 29, 1957), K. C. Bottenberg (to Phillips Petroleum Co.).

65. **U.S. Pat. 2,792,432** (May 14, 1957), W. C. Muller and F. D. Miller (to National Petroleum Chemicals Corp.).

66. **U.S. Pat. 2,792,433** (May 14, 1957), W. C. Muller and J. S. Atwood (to National Petro-Chemicals Corp.).

67. **U.S. Pat. 2,302,825** (Nov. 24, 1942), C. B. Wilde (to Stauffer Chemical Co.).

68. **U.S. Pat. 2,061,810** (Nov. 14, 1937), W. H. Shiffler and M. M. Holm (to Standard Oil of California).

69. **U.S. Pat. 2,038,512** (Apr. 21, 1936), R. N. Graham (to Union Carbide Corp.).

70. V. A. Gutuiyya and co-workers, *Azerb. Neft. Khoz.* **16**(5), 72 (1936).

71. **U.S. Pat. 2,474,568** (June 18, 1949), L. A. Bannon and C. E. Morrell (to Standard Oil Development Co.).

72. **U.S. Pat. 2,474,569** (June 18, 1949), L. A. Bannon (to Standard Oil Development Co.).

73. R. Kremann, *Monatsh. Chem.* **31**, 245, 275 (1910); R. Kremann, *J. Soc. Chem. Ind. (London)* **29**, 782 (1910).

74. R. Kremann, *Monatsh. Chem.* **28**(3), 13 (1907); **31**, 165 (1910); **38**, 53 (1917).

75. W. A. Drushei and G. A. Linhart, *Am. J. Sci.* **32**, 51 (1911).

76. **U.S. Pat. 2,414,759** (Jan. 21, 1947), H. O. Mottern (to Standard Alcohol Co.).

77. **U.S. Pat. 2,512,327** (June 20, 1950), T. P. Hawes and A. P. Geraitis (to Standard Oil Development Co.).

78. **Can. Pat. 470,593** (Jan. 2, 1951), T. P. Hawes and co-workers (to Standard Oil Development Co.).

79. **U.S. Pat. 2,856,265** (Oct. 15, 1958), G. A. Lescisin (to Union Carbide Corp.).

80. **Brit. Pat. 638,547** (June 17, 1950), J. Howlett and co-workers (to Distillers Co., Ltd.).

81. T. C. Carle and D. M. Stewart, *Chem. Ind. (London)* **19**, 830 (1962).

82. U.S. Pat. **41,685** (1861), E. A. Cotelte.

83. J. P. Wibaut and J. J. Dickman, *Proc. K. Ned. Akad. Wet.* **26**, 321 (1923).

84. C. R. Nelson and co-workers, *Chem. Eng. Progr.* **50**, 562 (1954).

85. U.S. Pat. **2,960,477** (Nov. 15, 1958), W. C. Smith, C. A. MacMurray, and W. L. Holmes (to Shell Oil Co.).

86. Ger. Pat. **2,015,536** (Nov. 5, 1970), O. Frampton and J. Feldman (to National Distillers and Chemical Corp.).

87. T. C. Carle and co-workers, *Chem. Ind.* **19**, 830 (1962).

88. W. J. Tapp, *Ind. Eng. Chem.* **40**, 1619 (1948); **42**, 1698 (1950); **44**, 2020 (1952).

89. Brit. Pat. **308,859** (1929), R. E. Slade (to ICI).

90. Brit. Pat. **378,865** (1932), L. H. Horsley (to ICI).

91. U.S. Pat. **2,579,601** (Dec. 25, 1951), C. R. Nelson and co-workers (to Shell Development Co.).

92. Brit. Pat. **1,144,947** (Mar. 12, 1969), H. G. Hagemeyer and M. Statman (to Eastman Kodak Co.).

93. U.S. Pat. **3,554,926** (Jan. 12, 1971), M. Statman and D. S. Martin (to Eastman Kodak Co.).

94. Brit. Pat. Appln. **15643** (1940) (to Standard Oil Development Co.).

95. U.S. Pat. **3,704,329** (Nov. 28, 1972), E. Rindtorff and W. Ester (to Veba-Chemnie Aktiengesellschaft).

96. U.S. Pat. **3,232,997** (Feb. 1, 1966), W. Ester (to Hibernia-Chemie Gesellschaft nit beschränkter Haftung).

97. M. A. Dalin, *Khem. Nauka i Promy* **1**, 259 (1956).

98. U.S. Pat. **4,012,452** (Mar. 15, 1977), O. D. Frampton (to National Distillers Corp.).

99. Jpn. Kokai **75,146,585** (Nov. 25, 1975), (to Nippon Synthetic Alcohol Co. Ltd.).

100. Jpn. Kokai **76,044,915** (Dec. 1, 1976) (to Nikki Chem. Co. Ltd.).

101. Brit. Pat. **394,376** (1933), H. Dreyfus.

102. U.S. Pat. **2,504,618** (Apr. 18, 1950), R. C. Archibald and R. A. Trimble (to Shell Development Co.).

103. J. Muller and H. I. Waterman, *Brennst. Chem.* **38**, 321 (1957).

104. U.S. Pat. **2,876,266** (Mar. 3, 1959), C. Wegner (to Farbenfabriken Bayer A. G.).

105. U.S. Pat. **2,769,847** (Nov. 6, 1956), R. L. Robinson (to Imperial Chemical Industries, Ltd.).

106. C. V. Mace and C. F. Bonilla, *Chem. Eng. Prog.* **50**, 385 (1954).

107. Brit. Pat. **691,360** (May 13, 1953), R. C. Thomson and R. K. Greenhaig (to Imperial Chemical Industries, Ltd.); Belg. Pat. **499,676** (May 28, 1951), R. C. Thomson and R. K. Greenhaig (to Imperial Chemical Industries, Ltd.).

108. U.S. Pat. **2,725,403** (Nov. 29, 1955), M. A. E. Hodgson (to Imperial Chemical Industries, Ltd.).

109. Brit. Pat. **665,214** (Jan. 16, 1952), J. Thompson and P. W. Reynolds (to Imperial Chemical Industries, Ltd.).

110. Winkler, *Ind. Agr. Aliment (Paris)* **66**, 159 (Mar., Apr. 1949).

111. U.S. Pat. **2,720,232** (Feb. 15, 1955), H. R. Arnold and J. E. Carnahan (to E. I. du Pont de Nemours & Co., Inc.).

112. U.S. Pat. **2,755,309** (July 17, 1956), P. W. Reynolds (to Imperial Chemical Industries, Ltd.).

113. U.S. Pat. **2,807,655** (Sept. 24, 1957), L. R. Pitwell (to Imperial Chemical Industries, Ltd.).

114. U.S. Pat. **2,763,697** (Sept. 18, 1956), H. J. Hagemeyer, Jr., and W. J. Clegg (to Eastman Kodak Co.).

115. Jpn. Kokai **5,1059,807** (May 25, 1976); **5,1063,105** (June 1, 1976); **5,1075,007** (June 19, 1976); **5,1086,405** (July 29, 1976) and **5,1141,802** (Dec. 7, 1976) (to Mitsui Toatsu Chem. Inc.).

116. Ger. Pat. **2,702,500** (July 28, 1977) (to Mitsui Toatsu Chem. Inc.).

117. F. C. Whitmore, *Ind. Eng. Chem.* **26**, 94 (1934).

118. W. S. Brey and K. A. Krieger, *J. Am. Chem. Soc.* **71**, 3637 (1949).

119. R. W. Taft, *J. Am. Chem. Soc.* **74**, 5372 (1952).

120. G. K. Boreskov and co-workers, *Tr. Nauchno. Issled. Inst. Spirtov Org. Prod.*, 213 (1960).

121. A. I. Gel'bshtein, Yu. M. Bakshi, and M. I. Temkin, *Dokl. Akad. Nauk SSSR* **132**, 384 (1960).

122. M. P. Applebey and co-workers, *J. Soc. Chem. Ind. Trans.* **56**, 279–81 (1937).

123. R. H. Bliss and B. F. Dodge, *Ind. Eng. Chem.* **29**, 19 (1937).

124. Yu. M. Bakshi, *Dokl. Akad. Nauk SSSR* **132**, 157 (1960).

125. J. Muller and H. I. Waterman, *Genie Chim.* **78**, 173 (1957).

126. Belg. Pat. **715,907** (Dec. 2, 1968) (to Hibernia-Chemie).

127. *Hydrocarbon Process.* **52**, 122 (1973).

128. Brit. Pat. **1,106,424** (Mar. 2, 1968) to Hibernia-Chemie.

129. U.S. Pat. **3,156,629** (Nov. 10, 1964), W. Ester (to Bergwerksgesellschaft Hibernia).

130. Brit. Pat. **1,007,096** (Oct. 13, 1965), E. Rindorff and co-workers (to Hibernia-Chemie).

131. Can. Pat. **859,781** (Dec. 29, 1970), D. C. Hull and G. I. LeMaster (to Eastman Kodak Co.).

132. U.S. Pat. **1,951,740** (March 20, 1934), W. H. Shiffler and M. M. Holm (to Standard Oil Development Co.).

133. U.S. Pat. **2,021,564** (Nov. 19, 1953), F. J. Metzger (to Air Reduction).

134. U.S. Pat. **1,607,459** (1927), O. Johannsen and O. Gross.

135. U.S. Pat. **2,179,092** (Nov. 7, 1939), V. N. Ipatieff (to Universal Oil Products).

136. Ger. Pat. **1,035,632** (Aug. 7, 1958), H. G. van Raay (to Farbwerke Hoechst AG).

137. *Chem. Weeks* **84**(6), 58 (1959).

138. Brit. Pat. **446,781** (1936), to Standard Oil Development.

139. E. R. Gilliland and co-workers, *Ind. Eng. Chem.* **38**, 370 (1936).

140. W. H. Shiffler, M. M. Holm, and L. F. Brooke, *Ind. Eng. Chem.* **31**, 1099 (1939).

141. N. S. Kozlov and N. Golubovskaya, *J. Gen. Chem. USSR* (Eng. Transl.) **6**, 1506 (1936).

142. U.S. Pat. **2,044,417** (June 16, 1936), F. R. Balcar (to Air Reduction).

143. Brit. Pat. **511,247** (1960), H. M. Guinot (to Usines de Melle).

144. Brit. Pat. **750,176** (1956) (to Chemische Werke Huls).

145. A. Vansheidt and M. Lozovskii, *Zh. Obshch. Khim.* **3**, 329 (1933); *Chem. Abstr.* **28**, 2323 (1934).

146. U.S. Pat. **3,095,458** (June 25, 1963), C. A. Judice and L. E. Pirkle (to Esso Research and Eng. Co.).

147. U.S. Pat. **2,045,785** (June 30, 1936), W. K. Lewis (to Standard Oil Development Co.).

148. U.S. Pat. **2,519,061** (Aug. 15, 1950), R. B. Mason (to Standard Oil Development Co.).

149. U.S. Pat. **2,115,874** (May 3, 1938), C. W. Rehm (to Union Carbide and Carbon).

150. T. V. Antipina and O. V. Isaer, *Zh. Fiz. Khim.* **31**, 2245 (1957); *Chem. Abstr.* **52**, 9944 (1958).

151. N. Ya. Kagan and co-workers, *J. Gen. Chem. USSR* (Eng. Transl.) **3**, 337 (1933).

152. U.S. Pat. **2,974,175** (Mar. 7, 1961), R. N. Watts and R. B. Mason (to Esso Research and Engineering Co.).

153. U.S. Pat. **2,472,618** (June 7, 1949), A. S. Ramage (to A. A. F. Maxwell).

154. U.S. Pat. **2,317,949** (Apr. 27, 1963), R. E. Burke (to Standard Oil Co. of Ohio).

155. W. Traube and R. Justh, *Brennst. Chem.* **4**, 150 (1923).

156. P. Sabatier and J. B. Senderens, *Comp. Rend.* **137**, 301 (1903).

157. S. A. Miller, ed., *Ethylene and Its Industrial Derivatives*, Ernest Benn Ltd., London, 1969.

158. D. M. Newitt and A. M. Block, *Proc. Royal Soc. (London)* **A140**, 426 (1933).

159. D. M. Newitt and T. T. A. Townsend, *J. Inst. Pet Technol.* **20**, 252A (1934).

160. P. J. Weizevich and P. K. Frolich, *Ind. Eng. Chem.* **26**, 267 (1934).

161. **U.S. Pat. 1,858,822** (May 17, 1932), P. K. Frolich (to Standard Oil Development Co.).

162. **U.S. Pat. 2,700,677** (Jan. 25, 1955), K. D. Bowen, D. R. Keck, and D. C. Lee, Jr. (to Celanese Corp. of America).

163. **U.S. Pat. 2,702,298** (Feb. 15, 1955), N. M. Caruthers (to Stanolind Oil and Gas Co.).

164. **U.S. Pat. 3,052,731** (Sept. 4, 1962), C. R. Murphy (to Gulf Research and Development Co.).

165. F. Fischer, *Ind. Eng. Chem.* **17**, 576 (1925).

166. R. Taylor and G. T. Morgan, *Proc. Roy. Soc.* **A131**, 533 (1931).

167. R. S. Aries, *Chem. Eng. News* **25**, 1792 (1947).

168. C. R. Downs and J. H. Rushton, *Chem. Eng. Prog.* **1**(1), 12 (1947).

169. P. C. Keith, *Am. Gas. Assoc. Mon.* **28**, 253 (1946).

170. M. D. Schlesinger and co-workers, *Ind. Eng. Chem.* **46**, 1322 (1954).

171. R. B. Anderson, J. Feldman, and H. H. Storch, *Ind. Eng. Chem.* **44**, 2418 (1952).

172. V. V. Kamzolin, *Priroda* **45**(11), 93 (1956).

173. I. Wender, R. A. Friedel, and M. Orchin, *Science* **113**, 206 (1951).

174. **Fr. Pat. 1,006,012** (April 18, 1953), H. Grasshof.

175. C. E. Morrell and co-workers, *Ind. Eng. Chem.* **44**, 2839 (1952).

176. *Pet. Refiner* **36**(10), 241 (1957).

177. G. T. Morgan and R. Tsylor, *Trans. Chem. Eng. Congr. London* **3**, 441 (1936).

178. **Belg. Pat. 842,430** (Dec. 1, 1976), L. H. Slaugh (to Shell Intl. Res.).

179. **Jpn. Kokai 5,2073,804** (June 21, 1977), (to Mitsubishi Gas Chem. Ind.).

180. **U.S. Pat. 4,062,898** (Dec. 13, 1977), M. Dubcek and G. C. Knapp (to Ethyl Corp.).

181. **Neth. Pat. 7,606,138** (1977), L. Slaugh (to Shell).

182. *Chem Eng. (New York)*, 18 (April 15, 1985).

183. *Hydrocarbon Process*, 13 (May 1985).

184. R. G. Herman, *Stud. Surf. Sci. Catal.* **64**, 265–349 (1991).

185. B. Cornils, *Crit. Rep. Appl. Chem.* **14**, 93–111 (1987).

186. X. Xu and co-workers, *Catal. Today* **2**(1), 125–170 (1987).

187. P. Courty and P. Chaumette, *Energy Prog.* **7**(1), 23–30 (1987).

188. H. B. M. Olayan, *Energy Prog.* **7**(1), 9–17 (1987).

189. *Chem. Eng. News*, 19 (Oct. 27, 1980).

190. *Chem. Mkt. Rep.* **3** (Aug. 12, 1991).

191. *Chem. Mkt. Rep.* **4**, 17 (July 30, 1973).

192. *Industrial Alcohol*, Misc. Publ. No. 695, U.S. Department of Agriculture, Feb. 1950.

193. *Chem. Eng. News* **29**(47), 4932 (1951).

194. E. S. Lepinsky, *Sugar J.*, 27 (Aug. 1976).

195. *A National Plan for Energy Research, Development, and Demonstration: Creating Energy Choices for the Future*, Vol. 1 ERDA, (1976).

196. *National Energy Outlook*, Federal Energy Administration, Washington, D.C., 1976.

197. *Report of Petrochemical Committee*, Government of India; *J. Indus. & Trade* **12**, 11 (1962).

198. N. Coutinho, *Bras. Acucareiro*, 19 (1976).

199. E. A. Jackson, *Process Biochem.* **29**, (1976).

200. *Bus. Week.*, 841 (Nov. 7, 1977).

201. S. K. Chan, *Tapioca*, Investigations at the Federal Experiment Station, Serdang, Malaysia, 1969.

202. W. A. Scheller and B. J. Mohr, *Chemtech* **7**(10), 616 (Oct. 1977).

203. W. A. Scheller, *Proceedings of the 8th National Conference on Wheat Utilization and Research*, USDA Pub. ARS-W19, Washington, D.C., Sept., 1974.

204. *Oil & Gas J.* **88**(41), 35 (Oct. 8, 1990).

205. *A Guide to Commercial-Scale Ethanol Production and Financing*, Report No. SERI SP-751-877, Solar Energy Research Institute, Golden, Colo., Mar. 1981.

206. Kane, S and LeBlanc M., *Ethanol and U.S. Agriculture*, Agriculture Information Bulletin Number 559, USDA, Washington, D.C., Jan. 1989.

207. *Chem. Mkt. Rep.* **240**(4), 3 (Jul. 22, 1991).

208. *Chem. Week.*, 17 (May 30, 1984).

209. *Chem. Eng. News* **54**(16), 12 (Apr. 12, 1976).

210. *Chemtech*, 315–318 (May 1980).

211. *Chem. Week.* **122**(14), 40 (Apr. 5, 1978).

212. *Chem. Week.* **149**(22), 36 (Jan. 1, 1992).

213. *Chem. Week.* **122**(12), 31 (Mar. 22, 1978).

214. *Environ. Sci. Technol.* **24**(12), 1765 (Dec. 1990), and **25**(2), 813 (May 1991).

215. *Chem. Mkt. Rep.*, **249**(9), 23 (Aug. 26, 1991).

216. L. Spano and co-workers, *29th Canadian Chemical Engineering Conference*, Oct. 1979.

217. Ref. 1, Chapt. 3, pp. 33–38.

218. **U.S. Pat. reissue 23,507** (May 27, 1952), R. L. Maycock and co-workers (to Shell Development Co.).

219. **U.S. Pat. 2,773,910** (Dec. 11, 1956), R. J. Schrader (to Eastman Kodak Co.).

220. **U.S. Pat. 2,981,661** (April 25, 1961), W. Sisco and J. S. Wiederecht (to American Cyanamid Co.).

221. **U.S. Pat. 3,014,971** (Dec. 26, 1961), S. W. Wilson (to Esso Research and Engineering Co.).

222. **U.S. Pat. 2,892,874** (June 30, 1959), C. J. B. Duculot.

223. **U.S. Pat. 2,721,874** (Oct. 25, 1955), A. McIlroy (to Stanolind Oil and Gas Co.).

224. **U.S. Pat. 2,892,757** (June 30, 1959), A. E. Markham (to Puget Sound Pulp and Timber Co.).

225. **U.S. Pat. 2,944,087** (July 5, 1960), E. W. Mommensen and co-workers (to Esso Research and Engineering Co.).

226. **U.S. Pat. 2,857,436** (Oct. 21, 1958), J. R. Mackinder and co-workers (to Shell Development Co.).

227. **U.S. Pat. 2,957,023** (Oct. 18, 1960), W. A. Dimler, Jr. (to Esso Research and Engineering Co.).

228. **U.S. Pat. 2,867,651** (Jan. 6, 1959), R. H. Wise (to Standard Oil Co.).

229. **U.S. Pat. 2,885,446** (May 5, 1959), S. P. Sharp and A. Steitz, Jr. (to Pan American Petroleum Co.).

230. Lee, F. and Pahl, R., *Ind. Eng. Chem. Process. Dev.* **24**, 168–172 (1985).

231. **U.S. Pat. 4,409,406** (May 27, 1982), J. Feldman (to National Distillers and Chemical Corp.).

232. **U.S. Pat. 4,410,745** (May 3, 1982), J. G. Victor (to Institute of Gas Technology).

233. **U.S. Patent 4,349,415** (September 28, 1979), R. P. DeFilippi and E. Vivian (to Critical Fluid Systems, Inc.).

234. **U.S. Pat. 4,400,241** (July 20, 1981), D. G. Braithwaite and co-workers (to Improtect).

235. *J. Commerce*, 21B (May 6, 1985).
236. *Chem. Week* **137**, 43 (Sept. 25, 1985).
237. **U.S. Pat. 4,345,973** (Aug. 24, 1982), M. R. Ladisch and G. T. Tsao (to Purdue Research Foundation).
238. *J. Commerce*, 1A (May 9, 1991).
239. *Oil & Gas J.* **88**(41), 38 (Oct. 8, 1990).
240. *Chem. Mkt. Rep.* **240**(4), 3 (July 22, 1991).
241. *Chem. Mkt. Rep.* **239**(12), 42 (Mar. 25, 1991).
242. *Chem. Mkt. Rep.* **240**(4), 28 (July 22, 1991).
243. *Chem. Mkt. Rep.*, 3 (Mar. 16, 1992).
244. *Chem. Week* **149**(1), 22 (July 31, 1991).
245. *Chem. Mkt. Rep.* **231**(10), (Mar. 13, 1992).
246. *Chem. Mkt. Rep.* **240**(20), 15 (Nov. 15, 1991).
247. *Oil & Gas J.* **88**(41), 38 (Oct. 8, 1990).
248. *Chem. Eng. News* **63**, (Apr. 22, 1985).
249. *Fed. Regist.* **56**(147), 36165 (July 31, 1991).
250. *Chem. Eng. News*, **63**(16), 9–15 (Apr. 22, 1985).
251. *Chem. Week* **137**(15), 13 (Oct. 9, 1985).
252. *Chem. Week* **137**(23), 5A (Dec. 4, 1985).
253. *Chem. Mkt. Rep.* **235**(18), 3, 25 (May 1, 1989).
254. *Chem. Mkt. Rep.* **238**(7), 5 (Aug. 13, 1990).
255. *U.S. Pharmacopeia XX-NFXV*, Mack Publishing Co., Easton, Pa., 1980.
256. *Reagent Chemicals, ACS Specifications*, 7th ed., American Chemical Society, Washington, D.C., 1987.
257. *1977 Book of ASTM Standards, Part 29, D-891, Method C*, American Society for Testing and Materials, Philadelphia, Pa., 1977.
258. V. C. Mehlenbocher, in J. Mitchell, Jr., ed., *Organic Analysis*, Vol. 1, Interscience Publishers, Inc., New York, 1953, pp. 1–65.
259. D. M. Smith and W. M. D. Bryant, *J. Am. Chem. Soc.* **57**, 61 (1935).
260. W. T. Robinson and co-workers, *Anal. Chem.* **33**, 1030 (1961).
261. S. Siggia and co-workers, *Anal. Chem.* **33**, 900 (1961).
262. P. R. Ashurst, *J. Inst. Brew.* **69**, 457 (1963).
263. M. Stiller, *Anal. Chim. Acta.* **25**, 85 (1961).
264. G. Machata, *Mikrochim. Acta.*, 691 (1962).
265. K. D. Parker and co-workers, *Anal. Chem.* **34**, 1234 (1962).
266. A. R. Alha and V. Tamminen, *Am. Med. Exptl. et Biol. Fenniae (Helsinki)* **38**, 121 (1960); *Chem. Abstr.* **55**, 13528 (1961).
267. H. V. Malmstadt and T. P. Hadjuannou, *Anal. Chem.* **34**, 455 (1962).
268. H. Croon and W. Croon, *Z. Anal. Chem.* **192**, 378 (1963).
269. P. Jaulmes and J. P. Laval, *Trau. Soc. Pharm. Montpellier* **21**, 21 (1962).
270. D. M. Comberbach and co-workers, *Chromatogr. Newsl.* **10**, 19 (1982).
271. B. B. Coldwell and C. L. Grant, *J. Forensic Sci.* **8**(2), 149 (1963).
272. K. Grosskopf, *Draeger Heft* **248**, 16 (1962); *Chem. Abstr.* **59**, 1023 (1963).
273. C. H. Thienes and T. J. Haley, *Clinical Toxicology*, 4th ed., Lea and Febiger, Philadelphia, 1964.
274. F. A. Patty, ed., *Industrial Hygiene and Toxicology*, Vol. 2, 2nd ed., Wiley-Interscience, New York, 1962.
275. N. I. Sax, *Dangerous Properties of Industrial Materials*, 4th ed., Van Nostrand Reinhold Co., Inc., New York, 1975.
276. L. T. Fairhall, *Industrial Toxicology*, 2nd ed., Williams and Wilkins Co., Baltimore, 1957.
277. R. B. Forney and F. W. Hughes, *Clin. Pharmacol. Ther.* **4**, 619 (1963).
278. H. W. Haggard and L. A. Greenberg, *J. Pharmacol. Exp. Ther.* **52**, 150 (1934).
279. E. J. Fredericks and M. Z. Lazor, *Am. Intern. Med.* **59**, 90 (1963).
280. R. E. Gosselin and co-workers, *Clinical Toxicology of Common Products, Acute Poisoning*, 4th ed., Williams and Wilkins Co., Baltimore, 1976.
281. A. J. Lanza and J. A. Goldberg, *Industrial Hygiene*, Thomas, Springfield, Illinois, 1939, p. 466.
282. E. Browning, *Toxicity and Metabolism of Industrial Solvents*, Elsevier, Amsterdam, The Netherlands, 1965, p. 326.
283. *Chem. Mkt. Rep.* **239**(12), 7 (Mar. 25, 1991).
284. *Code of Fed. Reg.*, 27 21.31–21.81 (Apr. 1, 1990).
285. *Statistical Release, Distilled Spirits*, U.S. Treasury Department, Bureau of Alcohol, Tobacco and Firearms, ATF, preliminary data for 1974–1975.
286. **U.S. Pat. 2,870,866** (Jan. 27, 1959), G. Baecklund (to Aktiebolaget Chematur).
287. **U.S. Pat. 3,073,752** (Jan. 15, 1963), M. Mention (to Les Usines de Melle SA).
288. **U.S. Pat. 3,154,586** (Oct. 28, 1964), O. E. Bander and co-workers (to Farbwerke Hoechst AG).
289. **U.S. Pat. 3,149,167** (Sept. 15, 1964), L. Hornig and co-workers (to Farbwerke Hoechst AG).
290. *Chem. Eng. News* **39**, 52 (Apr. 17, 1961).
291. **U.S. Pat. 3,172,913** (March 9, 1965), L. Hornig and co-workers (to Farbwerke Hoechst AG).
292. **U.S. Pat. 3,154,586** (Oct. 17, 1964), O. E. Bander and co-workers (to Farbwerke Hoechst AG).
293. **U.S. Pat. 3,122,586** (Feb. 25, 1964), W. Berndt (to Consortium für Elektrochemische Ind. GmbH).
294. **U.S. Pat. 3,106,579** (Oct. 8, 1963), L. Hornig and H. Lenzmann (to Farbwerke Hoechst AG).
295. **U.S. Pat. 3,119,874** (Jan. 28, 1964), E. Paszthory and W. Reimenschneider (to Farbwerke Hoechst AG).
296. **U.S. Pat. 3,118,001** (Jan. 14, 1964), W. Reimenschneider (to Farbwerke Hoechst AG).
297. **U.S. Pat. 3,104,263** (Sept. 17, 1963), W. Reimenschneider (to Farbwerke Hoechst AG).
298. **U.S. Pat. 3,119,875** (Jan. 28, 1964), A. Steinmetz and co-workers (to Farbwerke Hoechst AG).
299. **U.S. Pat. 3,087,968** (April 30, 1958), L. Hornig and co-workers (to Farbwerke Hoechst AG).
300. **U.S. Pat. 3,086,994** (April 23, 1963), J. Smidt and co-workers (to Consortium für Elektrochemische Ind. GmbH).
301. **U.S. Pat. 3,080,425** (March 5, 1963), J. Smidt and co-workers (to Consortium für Elektrochemische Ind. GmbH).
302. **U.S. Pat. 3,086,052** (April 16, 1963), J. Smidt and co-workers (to Consortium für Elektrochemische Ind. GmbH).
303. **U.S. Pat. 3,131,223** (April 23, 1964), J. Smidt and co-workers (to Consortium für Elektrochemische Ind. GmbH).
304. **U.S. Pat. 2,974,173** (March 7, 1961), R. B. Long and co-workers (to Esso Research and Engineering).
305. **U.S. Pat. 3,106,581** (Oct. 8, 1963), S. D. Neely (to Eastman Kodak Co.).
306. **U.S. Pat. 3,080,426** (March 5, 1963), I. Kirshenbaum, E. M. Amir, and E. J. Inchalik (to Esso Research and Engineering Co.).
307. C. F. Cullis and E. J. Newitt, *Proc. R. Soc. (London)* **A237**, 530 (1956); **A242**, 516 (1957).
308. J. Klassem and R. S. Kirk, *AIChE J.* **1**, 488 (1955).
309. **U.S. Pat. 1,977,750** (Oct. 23, 1935), C. O. Young (to Union Carbide Corp.).
310. A. Bielanski and co-workers, *Bull. Acad. Polon. Sci. Cl.* **34**, 533 (1956).
311. A. Bielanski and co-workers, *Bull. Acad. Polon. Sci. Cl.* **33**, 497 (1955).
312. A. A. Balandin and P. Titeni, *Dokl. Akad. Nauk SSSR* **113**, 1019 (1957).

313. U.S. Pat. **2,861,106** (Nov. 18, 1958), W. Optiz and W. Urbanski (to Knapsack-Giesheim A.G.).

314. U.S. Pat. **2,884,460** (April 28, 1959), V. I. Komarensky (to Heavy Minerals Co.).

315. A. Kh. Bork and O. A. Markova, *Zh. Fiz. Khim.* **22**, 1381 (1948).

316. S. Abe, Sci. *Papers Inst. Phys. Chem. Res. (Tokyo)* **40**, 331 (1943).

317. Ital. Pat. **509,424** (June 14, 1955), V. Martello and S. Cecotti (to Bombrini Parodi-Defino Societa per Azioni).

318. V. M. Nikitin, *J. Cen. Chem. USSR* (Eng. Transl.) **15**, 273 (1945).

319. V. M. Nikitin and N. V. Razumov, *J. Cen. Chem. USSR* (Eng. Transl.) **11**, 133 (1941).

320. K. V. Topchieva and K. Yun-Pin, *Dokl. Akad. Nauk SSSR* **101**, 305 (1955).

321. S. B. Ansimov and G. I. Khaidarov, *Zh. Obshch. Khim.* **18**, 40 (1948).

322. Z. E. Kosalapov, *J. Cen. Chem. USSR* (Eng. Transl.) **5**, 307 (1935).

323. U. Tsao and J. W. Reilly, *Hydrocarbon Proc.* **57**, 133 (Feb. 1978).

324. A. B. Metzner and co-workers, *AIChE J.* **5**, 496–501 (1959).

325. R. E. Parker and co-workers, *Chem. Rev.* **59**, 737 (1959).

326. B. Weibull and co-workers, *Acta. Chem. Scand.* **8**, 847 (1954).

327. R. M. Laird and co-workers, *J. Chem. Soc.* **B9**, 1062 (1969).

328. T. P. Hilditch, *Catalytic Processes in Applied Chemistry*, Chapman and Hall Ltd., London, 1929, p. 264.

329. *Synthetic Ethyl Alcohol in the Manufacture of Vinegar*, Tech. Bull. No. 19, Enjay Chemical Company.

330. U.S. Pat. **2,085,785** (July 6, 1937), R. R. Bottoms (to the Girdler Corp.).

331. Brit. Pat. **399,201** (Oct. 2, 1933) (to Rohm & Haas Co.).

332. Fr. Pat. **767,771** (July 24, 1934) (to I. G. Farbenindustrial A.G.).

333. U.S. Pat. **2,964,558** (Dec. 13, 1960), J. M. F. Leathers (to Dow Chemical Co.).

334. Neth. Pat. **134,896** (April 17, 1972) (to Distillers Co.).

335. Jap. Pat. **47-37404** (Sept. 20, 1972) (to Nippon Shokubai Kapaku Kogyo).

336. U.S. Pat. **3,458,561** (July 29, 1969), C. T. Kautter and co-workers (to Rohm & Haas Co.).

337. Brit. Pat. **1,257,371** (Dec. 15, 1971) (to Toyo Soda Manufac.).

338. U.S. Pat. **3,442,935** (May 6, 1969), L. A. Pine and co-workers (to Esso Res. and Eng.).

339. U.S. Pat. **3,457,299** (July 22, 1969), R. D. Closson and co-workers (to Ethyl Corp.).

340. Ref. 157, p. 792.

341. B. T. Brooks, *Ind. Eng. Chem.* **27**, 282 (1935).

342. U.S. Pats. **1,873,536** and **1,873,537** (Oct. 23, 1932), R. L. Brown and W. W. Odel.

343. I. E. Adadurov and P. Ya. Krainii, *J. Phys. Chem. USSR* **5**, 136 (1934).

344. J. B. Sendrens, *Ann. Chim. Phys.* **25**, 449 (1912).

345. R. N. Pease and C. C. Yung, *J. Am. Chem. Soc.* **46**, 2397 (1924).

346. Ibid., 390 (1924).

347. T. V. Antipina and O. V. Isaev, *Zh. Fiz. Khim.* **31**, 2078 (1957).

348. T. V. Antipina and M. D. Sinitsyna, *Vestn. Mosk. Univ., Ser. Mat. Mekh. Astron., Fiz. Khim.* **12**, 137 (1957).

349. Brit. Pat. **332,756** (July 31, 1930) (to N. V. de Bataasche Petr. Maatschappij).

350. Brit. Pat. **350,010** (June 5, 1931) (to N. V. de Bataasche Petr. Maatschappij).

351. C. E. Schildknecht, A. O. Zoss, and C. McKinley, *Ind. Eng. Chem.* **39**, 180 (1947).

352. W. G. Hanford and D. L. Fuller, *Ind. Eng. Chem.* **40**, 1171 (1948).

353. U.S. Pat. **1,959,927** (May 22, 1934), W. Reppe (to I. G. Farbenindustrie A.G.).

354. U.S. Pat. **2,021,869** (Nov. 19, 1935), W. Reppe and W. Wolff (to I. G. Farbenindustrie A.G.).

355. U.S. Pat. **2,066,076** (Dec. 19, 1936), W. Reppe and W. Wolff (to I. G. Farbenindustrie A.G.).

356. *Chem. Mark. Rep.* **241**(1), 27 (Jan. 6, 1992).

357. M. P. E. Berthelot and co-workers, *Am. Chim. Phys.* **65**, 385 (1861).

358. D. I. Saletan and R. R. White, *Chem. Eng. Prog. Symp. Ser.* **48**(4), 59 (1952).

359. USSR Pat. **145,235** (1962), A. A. Skladner and co-workers.

360. Belg. Pat. **613,426** (1962) (to Wacker-Chemie).

361. D. B. Keyes, *2nd Eng. Chem.* **24**, 1096 (1932).

362. E. Chadwick, *Ind. Chem.* **39**, 345 (1963).

363. P. K. Frohlich and co-workers, *J. Am. Chem. Soc.* **52**, 1565 (1930).

364. H. F. Hoering and co-workers, *Ind. Eng. Chem.* **35**, 575 (1943).

365. G. Edgar and W. H. Schuyler, *J. Am. Chem. Soc.* **46**, 64 (1924).

366. H. Essex and J. D. Clark, *J. Am. Chem. Soc.* **54**, 1290 (1932).

367. A. Mailhe and F. de Godon, *Bull. Soc. Chem. Fr.* **29**, 101 (1921).

368. W. J. Knox and T. N. Burbridge, *J. Am. Chem. Soc.* **65**, 999 (1963).

369. K. Kerby and S. Swann, *Ind. Eng. Chem.* **32**, 1607 (1940).

370. V. E. Tischenko, *J. Russ. Phys. Chem. Soc.* **38**, 355 (1906).

371. W. C. Child and H. Adkins, *J. Am. Chem. Soc.* **45**, 3013 (1923).

372. East Ger. Pat. **10,646** (1955), P. Herte and A. Loventz.

373. R. D. Morin and A. E. Bearse, *2nd Eng. Chem.* **43**, 1596 (1951).

374. U.S. Pat. **3,265,748** (1966), D. M. Hurt (to E. I. du Pont de Nemours & Co.).

375. A. W. Fleer and co-workers, *Ind. Eng. Chem.* **47**, 986 (1955).

376. Can. Pat. **321,646** (Apr. 19, 1932), G. Bloomfield and co-workers (to Commercial Solvents Corp.).

377. Brit. Pat. **338,518** (Nov. 19, 1930), H. Dreyfus.

378. Brit. Pat. **347,593** (Apr. 22, 1931) (to Holzverkohlungs-Industrie A.G.).

379. Brit. Pat. **359,430** (Oct. 14, 1931) (to Deutsche Gold- und Silber-Scheideanstalt vorm. Roessler).

380. E. Donath, *Chem. Ztg.* **12**, 1191 (1888); **51**, 924 (1927).

381. U.S. Pat. **1,663,350** (March 20, 1928), K. Roka (to Holzverkohlungs-Industrie A.G.).

382. Brit. Pat. **353,467** (July 16, 1931), H. F. Oxley and co-workers (to British Celanese Co., Ltd.).

383. St. Bakowski and L. Stepniewski, *Przem. Chem.* **20**, 142 (1936).

384. B. A. Bolotov and co-workers, *Zh. Prikl. Khim.* **28**, 299 (1955).

385. B. A. Bolotov and K. P. Katkova, *Zh. Prikl. Khim.* **28**, 414 (1955).

386. M. Guerbet, *Compt. Rend.* **128**, 511 (1899).

387. U.S. Pat. **2,457,866** (Jan. 4, 1949), C. A. Carter (to Union Carbide Corp.).

388. E. F. Pratt and D. G. Kubler, *J. Am. Chem. Soc.* **76**, 52 (1954).

389. J. Bolle and L. Bourgeois, *Compt. Rend.* **233**, 1466 (1951).

390. Brit. Pat. **655,864** (Aug. 1, 1951), M. Sulzbacher.

391. M. Sulzbacher, *J. Appl. Chem.* (London) **5**, 637 (1955).
392. **U.S. Pat. 2,762,847** (Sept. 11, 1956), R. E. Miller and G. E. Bennett (to Monsanto Chemical Co.).
393. I. I. Ostromuislenski, *J. Russ. Phys. Chem. Soc.* **47**, 1472 (1915).
394. W. M. Quattlebaum, W. J. Toussaint, and J. T. Dunn, *J. Amer. Chem. Soc.* **69**, 593 (1947).
395. **U.S. Pat. 2,421,361** (May 27, 1947), W. J. Toussaint and J. T. Dunn (to Union Carbide).
396. B. B. Corson and co-workers, *Ind. Eng. Chem.* **42**, 359 (1950).
397. I. R. Laszlo and co-workers, *Magy. Kem. Foly.* **60**, 65 (1944).
398. **U.S. Pat. 2,800,517** (July 23, 1957), C. Romanovsky and T. E. Jordan (to Publicker Industries, Inc.).
399. G. Natta and R. Rigamonti, *Chim. Ind. (Milan)* **29**, 239 (1947).
400. A. Brenman, *Rev. Chim. (Bucharest)* **8**, 286 (1957).
401. **U.S. Pat. 2,850,463** (Sept. 2, 1958), C. Romanovsky and T. E. Jordan (to Publicker Industries, Inc.).
402. M. Essayan and I. Ciolan, *Rev. Chim. (Bucharest)* **7**, 31 (1956).

John E. Logsdon
Union Carbide Corporation

ETHANOLAMINES.

See ALKANOLAMINES.

ETHERS

Ethers are compounds of the general formula Ar–O–Ar', Ar–O–R, and R–O–R' where Ar is an aryl group and R is an alkyl group. If the two R or Ar groups are identical, the compound is a symmetrical ether. Examples of symmetrical ethers are (di)methyl ether, CH₃OCH₃, and (di)phenyl ether, C₆H₅OC₆H₅; examples of unsymmetrical ethers are methyl ethyl ether, CH₃OCH₂CH₃, and methyl *tert*-butyl ether, CH₃OC(CH₃)₃. Cyclic ethers are variously designated. Three-membered ring cyclic ethers are olefin oxides, eg, ethylene oxide (1) (qv), 1,2-propylene oxide (qv), and 1,2-butylene oxide (see GLYCOLS). This ether group is also known as an epoxy group or oxirane. When the ring containing the ether is larger, the ethers are usually classified as oxygen heterocyclics such as tetrahydrofuran (THF) (2), or *p*-dioxane (3).



(1)



(2)



(3)

Simple ethers derive their name from the two groups attached to the oxygen followed by the word ether, eg, diethyl ether, CH₃CH₂OCH₂CH₃. For symmetrical ethers the "di" prefix is often omitted. If one group has no simple name, the compound may be named as an alkoxy derivative, eg, 2-ethoxyethanol, CH₃CH₂OCH₂CH₂OH.

Physical Properties

In general, ethers are neutral, pleasant-smelling compounds that have little or no solubility in water, but are easily soluble in organic liquids (1). Their boiling points approximate those of hydrocarbons having comparable molecular weights and geometries. Table 1 gives the basic physical properties for many of the ethers. More detailed physical properties for the ethers most commonly used as solvents are listed in Table 2.

Table 1. Physical Properties of Some Representative Ethers

Systematic name	Common name	CAS Registry Number	Bp, °C	<i>d</i> ₄ ²⁰ ^b	<i>n</i> _D ²⁰
<i>Saturated</i>					
symmetrical					
methyl ether		[115-10-6]	23.7	1.617 ^c	
2-methoxyethyl ether		[111-96-6]	162	0.9451	1.4097
ethyl ether		[60-29-7]	34.5	0.7146	1.3527
1-chloroethyl ether		[986-48-7]	116–117	1.1060 ²⁵	1.4185 ²⁵
<i>n</i> -propyl ether		[111-43-3]	90.5	0.7360	1.3809
isopropyl ether		[108-20-3]	68.5	0.7257	1.3682
<i>n</i> -butyl ether		[142-96-1]	142.0	0.7704	1.3981
<i>sec</i> -butyl ether		[6863-58-7]	122.0	0.7590 ²⁵	1.3931
isobutyl ether		[628-55-7]	123.0	0.7612 ¹⁵	
<i>tert</i> -butyl ether		[6163-66-2]	108.0	0.7622 ¹⁵	1.3946
<i>n</i> -amyl ether		[693-65-2]	188.0	0.7849	1.4119
isoamyl ether		[544-01-4]	173.0	0.7777	1.4085
<i>sec</i> -amyl ether		[56762-00-6]	161.0	0.7830	1.4058
<i>n</i> -hexyl ether		[112-58-3]	223 ₁₀₂	0.7936	1.4204
<i>n</i> -heptyl ether		[629-64-1]	258.5 ₁₀₂	0.8008	1.4275
<i>n</i> -octyl ether		[629-82-3]	286–287	0.8063	1.4327
unsymmetrical					
methyl <i>n</i> -propyl ether		[557-17-5]	38.9	0.738	1.3579
methyl isopropyl ether		[598-53-8]	32.5 ₁₀₄	0.7237 ¹⁵	1.3576
methyl <i>n</i> -butyl ether		[628-28-4]	70.5	0.7443	1.3736
methyl isobutyl ether		[625-44-5]	105–106 ₉	0.7549	1.3852 ²⁵

methyl <i>tert</i> -butyl ether	MTBE	[1634-04-4]	55.1	0.7406	1.3690
methyl <i>tert</i> -amyl ether	TAME	[994-05-8]	85–86	0.770	1.3896
ethyl isopropyl ether		[625-54-7]	53.5	0.7211	1.3624
ethyl <i>n</i> -butyl ether		[628-81-9]	91.5	0.7490	1.3818
ethyl <i>tert</i> -butyl ether	ETBE	[637-92-3]	72–73	0.742	1.3756
ethyl <i>n</i> -amyl ether		[17952-11-3]	118.0	0.7622	1.3927
ethyl <i>tert</i> -amyl ether		[919-94-8]	101	0.7657	1.3912
isopropyl <i>tert</i> -butyl ether		[17348-59-3]	87.6	0.7365 ²⁵	1.3799
2-ethoxyethanol	cellosolve	[110-80-5]	135	0.931 ²⁰	1.406 ²⁵
2-(2-ethoxy)ethoxyethanol ether	carbitol	[111-90-0]	196	0.9855 ²⁵	1.4273
<i>Unsaturated</i>					
vinyl ether		[109-93-3]	28–31	0.767 ²⁵	
vinyl methyl ether		[107-25-5]	5–6	0.7511	
vinyl ethyl ether		[109-92-2]	35.0	0.7533	1.3739 ²⁵
vinyl <i>n</i> -butyl ether		[111-34-2]	93.5	0.7735 ²⁵	1.3997 ²⁵
allyl ether		[557-40-4]	95.0	0.8053	1.4163
bis(2-methallyl) ether		[628-56-8]	105.4	0.8627	1.4206
allyl ethyl ether		[557-31-3]	67.6	0.765	1.3881
allyl glycidyl ether		[106-92-3]	75.0 ₇		1.4310 ³⁰
ethynyl ethyl ether		[927-80-0]	49.0	0.8001	1.3796
ethynyl butyl ether		[3329-56-4]	104.0	0.8078 ²⁵	1.4033 ²⁵
<i>Cyclic</i>					
ethylene oxide	oxirane	[75-21-8]	13.5 _{99.4}	0.8824 ¹⁰	1.3597 ⁷
1,2-propylene oxide	methyloxirane	[75-56-9]	34.3	0.8590 ¹⁰	1.3670
1,3-propylene oxide	oxetane	[503-30-0]	47.8	0.8930 ²⁵	1.3961
tetrahydrofuran	oxolane	[109-99-9]	64.5	0.8892	1.4050
furan	oxole	[110-00-9]	31.36	0.9514	1.4214 ²
tetrahydropyran	oxane	[142-68-7]	88	0.8810	1.4200
1,4-dioxane		[123-91-1]	101 ₁₀₀	1.0337	1.4224
<i>Aromatic</i>					
methyl phenyl ether		[100-66-3]	153.8	0.9954	1.5179
4-methoxytoluene		[104-93-8]	176.5	0.9689	1.5124
ethyl phenyl ether		[103-73-1]	172	0.9792	1.5076
	<i>trans</i> -anethole	[4180-23-8]	253	0.9882	1.5615
1-methoxy-4- <i>trans</i> -propenylbenzene			21 ^d		
	estragole	[140-67-0]	215	0.9645	1.5230
1-methoxy-4-allylbenzene phenyl		[101-84-8]	258	1.0863	1.5780
			28 ^d		
2-methoxyphenol	guaiacol	[90-05-1]	205.5	1.1287	1.5385
1,2-dimethoxybenzene	veratrole	[91-16-7]	206.7	1.084	1.5385
			22–23 ^d		
1,4-dimethoxybenzene		[150-78-7]	212.6	1.0526 ⁵⁵ ₅₅	
			58–60 ^d		
2-methoxy-4-allylphenol	4-allylguaiacol	[97-53-0]	255	1.0664	1.5410
1,2-dimethoxy-4-allylben-zene	4-allylveratrole	[93-15-2]	248	1.055	1.532
1-allyl-3,4-methylenedioxy-benzene	safrole	[94-59-7]	234.5	1.0950	1.5383
1-propenyl-3,4-methylene-dioxybenzene	isosaftrole	[93-16-3]	252	1.1224	1.5782
2-methoxy-4- <i>cis</i> -propenyl-phenol	<i>cis</i> -isoeugenol	[5912-86-7]	134 ₁₇	1.0851	1.5700
2-methoxy-4- <i>trans</i> -propenyl-phenol	<i>trans</i> -isoeugenol	[5932-68-3]	141 ₁	1.0852	1.5782
1-benzyl-2-methoxy-4- <i>trans</i> -propenylbenzene	benzyl isoeugenol	[120-11-6]	58 ^d		
			264–270 ₉₈		
		[25013-16-3, 55949-47-8] [121-00-6] [88-32-4]	48–55 ^d		
butyrated hydroxyanisole (BHA), a mixture of: 2- <i>tert</i> -butyl-4-methoxy-phenol and 3- <i>tert</i> -butyl-4-methoxyphenol					

^a At 101.3 kPa unless otherwise noted by subscript in kPa; to convert kPa to mm Hg, multiply by 7.5.

^b Unless otherwise noted as superscript, °C.

^c Specific gravity (air).

^d Melting point.

Table 2. Typical Physical and Chemical Properties of Commonly Used Ethers^a

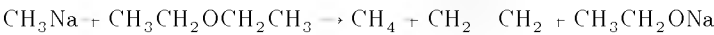
Property	Ethyl	MTBE	THF	Isopropyl	<i>n</i> -Butyl
chemical formula	C ₄ H ₁₀ O	C ₅ H ₁₂ O	C ₄ H ₈ O	C ₃ H ₈ O	C ₄ H ₁₀ O
molecular weight	74.12	88.15	72.10	102.17	139.22
boiling point at 101.3 kPa, ^b °C	34.5	55.0	66.0	68.4	142.0
vapor pressure at 20°C, kPa ^b	56	27	17	16	0.67
evaporation rate ^c	11.8	8.5	5.7	8.0	0.66
viscosity at 20°C, mPa·s(cP)	0.23	0.35	0.48	0.38	0.65
surface tension in air, 25°C, mN m	17.0	18.3	26.4	32.0	22.9

(dyn cm)					
dipole moment, 25–50°C, C·m ^d	4.3 × 10 ^{−30}	4.7 × 10 ^{−30}	5.3 × 10 ^{−30}	4.7 × 10 ^{−30}	4 × 10 ^{−30}
Hilderbrand solubility parameter	7.4	7.4	9.1	7.1	7.7
water solubility, wt % , 20°C ^e					
ether in water	6.9	4.8	inf.	1.07	0.03
water in ether	1.3	1.4	inf.	0.53	0.19
flash point	40	30	17	13	25
autoignition temp, °C	160	426	321	440	185
flammability limits in air, vol %					
lower	1.9	1.6	1.8	1.0	0.9
higher	48.0	8.4	11.8	21.0	8.5
reactivity with hydroxyl radical ^f	13.6	3.2	17.8		27.8

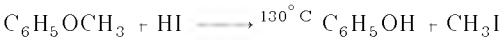
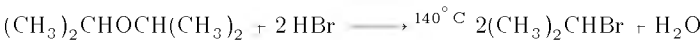
^a Ref. 2.
^b To convert kPa to mm Hg, multiply by 7.5.
^c *n*-Butyl acetate 1.
^d To convert C·C·m to debyes, multiply by 3 × 10²⁹.
^e Ref. 3.
^f *k*_{OH} (298 K), 10¹² cm³·mol^{−1}·sec^{−1}; Refs. 4 and 5.

Chemical Properties

Most ethers, particularly dialkyl ethers, are comparatively unreactive compounds because the carbon-oxygen bond is not readily cleaved. For this reason, ethers are frequently employed as inert solvents in organic synthesis. However, within the ether family, the cyclic ethers are more reactive, the most reactive being the olefin oxides or epoxides. Epoxides are generally used as intermediates for producing other small molecules or polymers (see EPOXY RESINS; POLYETHERS). Ethers do react with exceptionally powerful basic reagents, particularly certain alkali metal alkyls, to give cleavage products (6). Ethers react with less powerful bases to give the same cleavage products, but only under the forcing conditions of high temperature and pressure.

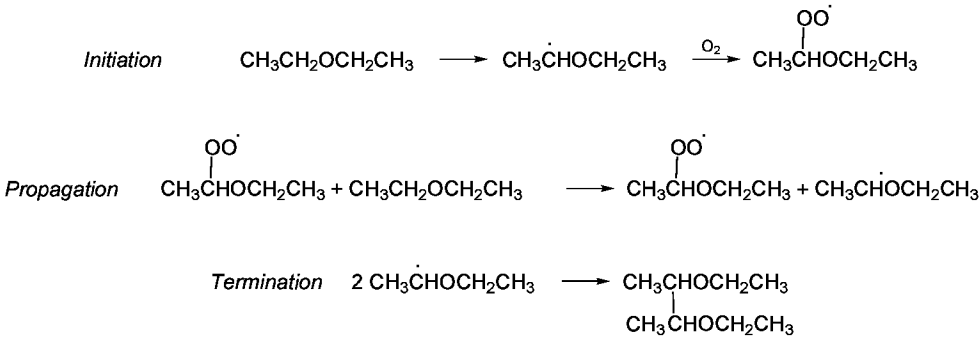


The ether linkage can also be cleaved by strong acids, generally at high temperatures (6):



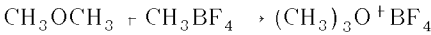
Other acids that have been used for this cleavage are phosphoric acid, pyridine hydrochloride, boron tribromide, trifluoroacetic acid, and nitric acid. Acid cleavage of aryl alkyl ethers always gives phenol because the aromatic carbon–oxygen bond is much stronger than the aliphatic carbon–oxygen bond. Unsymmetrical aliphatic ethers usually yield a mixture of alkyl halides and alcohols when cleaved by halogen containing acids.

Most ethers are potentially hazardous chemicals because, in the presence of atmospheric oxygen, a radical-chain process can occur, resulting in the formation of peroxides that are unstable, explosion-prone compounds (7). The reaction may be generalized in terms of the following steps involving initiation, propagation, and termination.



The nature of the initiation step, which may occur in a variety of ways, is not known in all cases. Commonly used ethers such as ethyl ether, isopropyl ether, tetrahydrofuran, and *p*-dioxane are particularly prone to form explosive peroxides on prolonged storage and exposure to air and light (see PEROXIDES AND PEROXY COMPOUNDS, ORGANIC), and should contain antioxidants (qv) to prevent their build-up. One of the exceptions to the peroxide forming tendency of ethers is methyl *tert*-alkyl ethers such as methyl *tert*-butyl ether [1634-04-4] (MTBE) and *tert*-amyl methyl ether [994-05-8] (TAME). Both have shown little tendency if any to form peroxides (2,8).

Ethers are weakly basic and are converted to unstable oxonium salts by strong acids such as sulfuric acid, perchloric acid, and hydrobromic acid; relatively stable complexes are formed between ethers and Lewis acids such as boron trifluoride, aluminum chloride, and Grignard reagents (qv) (9):



Like other aromatic compounds, aromatic ethers can undergo substitution in the aromatic ring with electrophilic reagents, eg, nitration, halogenation, and sulfonation. They also undergo Friedel-Crafts (qv) alkylation and acylation.

Allyl phenyl ethers rearrange cleanly at high temperatures, producing *o*-allyl phenols or *p*-allylphenols if both ortho positions are blocked. This reaction is called the Claisen rearrangement (10).

Fuel Properties

In addition to MTBE, two other ethers commonly used as fuel additives are *tert*-amyl methyl ether (TAME) and ethyl *tert*-butyl ether [637-92-3] (ETBE). There are a number of properties that are important in gasoline blending (see GASOLINE AND OTHER MOTOR FUELS) (Table 3).

Table 3. Typical Gasoline-Related Properties for Ethers Used in Gasoline^a

Property	MTBE	ETBE	TAME
----------	------	------	------

research octane number (RON)	118	119	112
motor octane number (MON)	102	104	99
(RON + MON) / 2	110	111	105.5
boiling point			
°F	131	161	187
°C	55	72	86
Reid vapor pressure, kPa ^b	55	28	7
oxygen, wt %	18.2	15.7	15.7

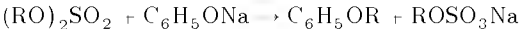
^a Ref. 11.
^b To convert kPa to psi, multiply by 0.145.

Preparation

The most versatile method of preparing ethers is the Williamson ether synthesis, particularly in the preparation of unsymmetrical alkyl ethers (12,13). The reaction of sodium alcoholates with halogen derivatives of hydrocarbons gives the ethers:

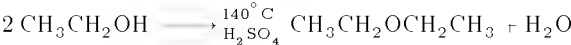


Dialkyl sulfates can replace the halogen derivatives, and this modification is especially useful for the preparation of phenolic ethers:



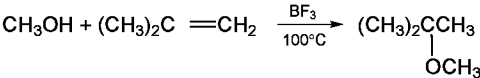
Aromatic halides do not react easily with phenoxide ions to produce diaryl ethers unless the aromatic halide is substituted with one or more electron-withdrawing groups, eg, nitro or carboxyl groups. The Ullmann reaction uses finely divided copper or copper salts to catalyze the reaction of phenoxides with aromatic halides to give diaryl ethers.

Alcohols can be dehydrated with strong acid catalysts and high reaction temperatures to produce ethers. This method is particularly useful for the preparation of symmetrical lower alkyl ethers, such as ethyl ether.



The reaction gives poor yields of ethers with secondary and tertiary alcohols; dehydration to form the corresponding olefin is a more favorable reaction. The reaction fails for the production of diaryl ethers from phenols.

Alkyl tertiary alkyl ethers can be prepared by the addition of an alcohol or phenol to a tertiary olefin under acid catalysis (Reylder reaction); sulfuric acid, phosphoric acid, hydrochloric acid, and boron trifluoride have all been used as catalysts:



Commercially, sulfonic acid ion-exchange resins are used in fixed-bed reactors to make these tertiary alkyl ethers (14). Since the reaction is very selective to tertiary olefins and also reversible, a two-step procedure is also used to recover commercially pure tertiary olefins from mixed olefin process streams. The corresponding tertiary alkyl ether is produced in the olefin mixture and then easily separated from the unreacted olefins by simple fractionation. The reaction is then reversed in a second step to make a commercially pure tertiary olefin, usually isobutylene or isoamylene.

Ether production from the reaction of monohydric alcohols with linear olefins has also been demonstrated under more severe conditions using a zeolite-type catalyst system (15). However, it has not been commercialized as yet.

Economic Aspects

Table 4 compares the prices and availability of several commercially available ethers as of January 1992.

Table 4. Price and Availability of Some Commercial Ethers^a

Ether	Price, \$ /kg	Availability
methyl (aerosol-grade)	0.95	tank car
ethyl	1.17	tank car
<i>n</i> -butyl	3.08	208-L ^b drum
tetrahydrofuran	2.77	tank car
isopropyl	1.01	tank car
methyl <i>tert</i> -butyl	0.99 ^c	tank truck
butylated hydroxy anisole (BHA) ^d	19.36	208-L ^b drum

^a Ref. 16.
^b 55-gal.
^c Refined-grade (99.9% purity).
Courtesy of ARCO Chemical Co.
^d A mixture of 2-*tert*-butyl-4-methoxyphenol and 3-*tert*-butyl-4-methoxyphenol.

Fuel-grade ethers are generally produced in crude oil refineries for internal use in the manufacture of gasoline or in petrochemical plants for resale to the refiners. The value of these ethers fluctuates with the price of gasoline and octane in the industry. It is not uncommon for the market price of MTBE to change weekly and sometimes daily. The two largest MTBE markets are in the U.S. Gulf Coast region and in Northern Europe. MTBE is typically sold to refiners in large barges in quantities of ten to one hundred thousand U.S. barrels (1 to 12 10³ t). It is also sometimes shipped in batches on gasoline product pipelines. Spot market prices quoted in January 1992 ranged from 92 to 96 U.S. cents per gallon (\$244–\$254 /m³) for the U.S. Gulf Coast region (17).

Health and Safety Factors

Although ethers are not particularly hazardous, their use involves risks of fire, toxic effects, and several unexpected reactions.
Flammability. Since almost all ethers burn in air, an assessment of their potential hazards depends on flash points and ignition temperatures. The flash point of a liquid is the lowest temperature at which vapors are given off in sufficient quantities for the vapor–air mixture above the surface of the

liquid to propagate a flame away from the source of ignition. In other words, an explosive vapor–air mixture can form whenever a liquid is used or stored in an open container at a temperature above its flash point. Table 2 includes the flash points of several common ethers.

The ignition temperature is the minimum temperature required to initiate or cause self-sustained combustion. Table 2 also lists ignition temperatures of several common ethers. Attention is directed to the particularly low ignition temperature of ethyl ether, especially with reference to some common ignition sources such as a lighted cigarette (732°C) or a pressurized (0.7 MPa or 100 psi) steam line (180°C).

Toxicity. The effect of ethers owing to ingestion, skin contact, or inhalation may range from drowsiness and lack of coordination to serious injury or death. Ethers are reported to have a low order of toxicity when ingested, although 30–60 mL may be fatal when swallowed (18), and a case of fatal poisoning owing to ingestion of a large quantity of 2-methoxyethanol has been recorded (19).

Prolonged or repeated contacts of ethers with skin cause tissue defatting and dehydration leading to dermatitis. Some compounds penetrate the skin in harmful amounts. 1,4-Dioxane and 1-allyl-3,4-methylenedioxybenzene (safrole) have been listed as Category I carcinogens by OSHA. Category I carcinogens are confirmed cancer-causing agents based on human data, or based on tests in two mammalian species, or in one species if the tests have been replicated.

Inhalation is the most common means by which ethers enter the body. The effects of various ethers may include narcosis, irritation of the nose, throat, and mucous membranes, and chronic or acute poisoning. In general, ethers are central nervous system depressants, eg, ethyl ether and vinyl ether are used as general anesthetics.

Peroxide Formation. Except for the methyl *tert*-alkyl ethers, most ethers tend to absorb and react with oxygen from the air to form unstable peroxides that may detonate with extreme violence when concentrated by evaporation or distillation, when combined with other compounds that give a detonable mixture, or when disturbed by heat, shock, or friction. Appreciable quantities of crystalline solids have been observed as gross evidence for the formation of peroxides, and peroxides may form a viscous liquid in the bottom of ether-filled containers. If viscous liquids or crystalline solids are observed in ethers, no further tests for the detection of peroxides are recommended. Several chemical and physical methods for detecting and estimating peroxide concentrations have been described. Most of the qualitative tests for peroxides are readily performed and strongly recommended when any doubt is present (20).

Applications

Alkyl ethers are used for organic reactions and extractions, as plasticizers, as vehicles for other products, as anesthetics (qv), and octane (and oxygen) enhancers in gasoline. Most ethers have very low solubility in water, but dissolve most organic compounds, and therefore have found wide application in paint and varnish removers; as high boiling solvents for gums, resins, and waxes; in lubricating oils (see LUBRICATION); as an extraction solvent in the fragrance industry; and as an inert reaction medium in the pharmaceutical industry. The vapors of certain ethers are toxic to insects and are useful as agricultural insecticides (see INSECT CONTROL TECHNOLOGY) and industrial fumigants. Except for those used in gasoline, the lower mol wt technical-grade ethers are the least expensive and most commonly used in industry.

Aryl ethers have distinctive, pleasant odors and flavors which make them valuable to the perfume (qv) and flavor industries (see FLAVORS AND SPICES). Because of their heat stability, they are useful as heat-transfer fluids (see HEAT EXCHANGE TECHNOLOGY). Other aryl ethers are useful as food preservatives and antioxidants (see FOOD ADDITIVES).

Commercially Important Ethers

ETHYL ETHER

Diethyl ether [60-29-7] is one of the more important members of the ether family. It is a colorless, very volatile, highly flammable liquid with a sweet, pungent odor and burning taste. As a commercial product it is available in several grades; it is used in chemical manufacture, as a solvent, extractant, or reaction medium, and as a general anesthetic.

Manufacture. Much of the diethyl ether manufactured is obtained as a by-product when ethanol (qv) is produced by the vapor-phase hydration of ethylene (qv) over a supported phosphoric acid catalyst. Such a process has the flexibility to adjust to some extent the relative amounts of ethanol and diethyl ether produced in order to meet existing market demands. Diethyl ether can be prepared directly to greater than 95° yield by the vapor-phase dehydration of ethanol in a fixed-bed reactor using an alumina catalyst (21).

Handling and Shipment. The handling of ethyl ether is hazardous because of its highly flammable properties. Not only is it highly volatile, but it also has a low autoignition temperature and, as a nonconductor, can generate static electrical charges that may result in ignition or vapor explosion. The area in which ethyl ether is handled should be considered a Class I hazardous location as defined by the National Electrical Code (22). All tools used in making connections or repairs should be of the nonsparking type. All possible care should be taken in loading and unloading tank cars. The tank cars should be properly spotted, and the usual caution signs and derails placed in position. Before any connection is made, the tank car should be grounded and bonded. Tank cars should always be unloaded through dome connections rather than through bottom outlets, eg, a pressure-type LPG tank car. A positive suction-type pump or natural gas can be used to remove the ether from the tank car. In no event should air pressure be used.

Special containers have been developed for anesthetic ether to prevent deterioration before use. Their effectiveness as stabilizers usually depends on the presence of a lower oxide of a metal having more than one oxidation state. Thus the sides and the bottoms of tin-plate containers are electroplated with copper, which contains a small amount of cuprous oxide. Stannous oxide is also used in the linings for tin containers. Instead of using special containers, iron wire or certain other metals and alloys or organic compounds have been added to ether to stabilize it.

Economic Aspects. The production of ethyl ether from 1956 through 1973 ranged from 29.5 to 48.6 × 10⁶ kg as reported by Synthetic Organic Chemicals, U.S. Production and Sales. Production was estimated at 13.6–18 × 10⁶ kg in 1986, 12.7 × 10⁶ kg in 1989. Though 1990 U.S. production capacity was estimated at 25.5 × 10⁶ kg, production was estimated as only 12 × 10⁶ kg in 1991 (21). Much of the decrease has been the result of a decline in arsenal demand (smokeless gun powder). List prices for ether have been steadily increasing, and reached \$1.12 /kg by 1989, refined, tanks (fob).

Analytical Methods. Most of the analytical and testing methods used for ethyl ether are conventional laboratory methods. Ethyl ether that is to be used for anesthetic purposes or in processes that involve heating or distillation must be peroxide-free, and should pass the USP standard test with potassium iodide. This test detects approximately 0.001° peroxide as hydrogen peroxide.

Specifications. Ethyl ether is commercially available in the following grades: USP anesthesia, absolute (ACS), industrial, solvent (conc), and synthetic. Specifications vary, depending on the consumer and use. In many instances, the ether has to meet a specific test written into the specification, eg, it may be important that the ether is completely anhydrous or free from alcohol and aldehyde.

The technical concentrated ether contains very small amounts of alcohol, water, aldehydes, peroxides, and other impurities (Table 5). The more refined grades, such as anesthetic ether, are obtained from technical ether by redistillation and dehydration followed by alkali or charcoal treatment.

Table 5. Specifications for Various Grades of Ethyl Ether

Specification	Technical refined	ACS, absolute	USP XIX ^a
color, max	water-white	water-white	colorless
d_{25}^{20}	0.710–0.713	0.7079	0.713–0.716
acidity as acetic acid, wt %	0.0025	0.0010	passes test ^b
peroxides, max wt %	passes test ^c	0.0001	passes test ^c
aldehydes, max wt %	passes test ^d	0.0005	passes test ^d
alcohol, max %	0.5 vol %	0.05 wt %	

nonvolatile matter, max wt %	0.002	0.0010	0.003
water, max wt %	0.3	0.0100	
odor	nonresidual	passes test ^c	passes test ^c
net container contents, kg ^f			
19-L (5-gal) drum	13.6	13.6	13.6
208-L (55-gal) drum	147	147	147

^a For use in anesthesia; the USP (22) also recognizes slightly less pure grades: ethyl oxide (solvent ether), ether abs, and reagent-grade ether.

^b Free acid, requiring no more than 0.4 mL of 0.02 *N* sodium hydroxide for 25 mL of ethyl ether.

^c No color with potassium iodide reagent; USP test.

^d No turbidity with alkaline mercuric chloride–potassium iodide reagent; USP test.

^e No foreign odor when last traces of ether evaporate from odorless absorbent paper; USP test.

^f The tech-grade can also be shipped in 103 W insulated tank cars of 15.1-m³ (4000-gal), 22.7-m³ (6000-gal), 30.3-m³ (8000-gal), and 37.9-m³ (10,000-gal) capacity.

Ethyl ether is classified by the ICC as a flammable liquid. As such it must be packed in ICC specification containers when shipped by rail, water, or highway; all ICC regulations regarding loading, handling, and labeling must be followed. Each container of ethyl ether must carry an identifying label or stencil. Tank cars and boxcars, either carload or less than carload, must bear the ICC dangerous placard. Each drum or each box with inside containers must bear the ICC red label for flammable liquids.

Toxicology. The toxicity of ethyl ether is low and its greatest hazards in industry are fire and explosion. The vapor is absorbed almost instantly from the lungs and very promptly from the intestinal tract. It undergoes no chemical change in the body. Prevention and control of health hazards associated with the handling of ethyl ether depend primarily on prevention of exposure to toxic atmospheric concentrations and scrupulous precautions to prevent explosion and fire.

A concentration of 35,000 ppm in air produces unconsciousness in 30–40 minutes. This concentration also constitutes a serious fire and explosion hazard, and should not be permitted to exist under any circumstance. Any person exposed to ethyl ether vapor of any appreciable concentration should be promptly removed from the area. Recovery from exposure to sublethal concentrations is rapid and generally complete. Except in emergencies, and then only with appropriate protective equipment, no one should enter an area containing ether vapor until the concentration has been found safe by measurement with a combustible-gas indicator.

If an ethyl ether fire occurs, carbon dioxide, carbon tetrachloride, and dry chemical fire extinguishers meeting National Fire Prevention Association Code 1 and 2 requirements may be used successfully (23). Water may also be effectively applied (see PLANT SAFETY). Hose streams played into open tanks of burning ethyl ether serve only to scatter the liquid and spread the fire. However, ether fires may be extinguished by a high pressure water spray that cools the burning surface and smothers the fire. Automatic sprinklers and deluge systems are also effective.

Uses. Ethyl ether [60-29-7] has a wide range of uses in the chemical industry. It is a good solvent or extractant for fats, waxes, oils, perfumes, resins, dyes, gums, and alkaloids. When mixed with ethanol, ethyl ether becomes an excellent solvent for cellulose nitrate in the manufacture of guncotton (see EXPLOSIVES AND PROPELLANTS), collodion solutions (see MEMBRANE TECHNOLOGY), and pyroxylin plastics (see CELLULOSE ESTERS). Another important use is an extractant for acetic acid as well as other organic acids, eg, in the cellulose acetate and plastic industries to recover acetic acid from dilute aqueous systems. Ethyl ether is also used as a denaturant in several denaturant alcohol formulas. It has been used as a starting fuel for diesel engines and as an entrainer for dehydration of ethanol and isopropyl alcohol. It may be used as an anhydrous, inert reaction medium for the Grignard and Wurtz-Fittig reactions. Ethyl ether is used as a general anesthetic in surgery.

The estimated use pattern of ethyl ether during 1990 was as follows: solvents and military production of smokeless powder, 35% ; chemical synthesis and solvent extraction, 35% ; diesel starting fluid, 30% (21).

METHYL TERTIARY BUTYL ETHER

MTBE was first commercially produced in Italy in 1973 for use as an octane enhancer in gasoline. U.S. production of MTBE started in 1979 after Atlantic Richfield Co. (ARCO) was granted a waiver by the U.S. Environmental Protection Agency (EPA) that allowed MTBE to be blended up to 7 vol % in U.S. unleaded gasoline. The use of other aliphatic ethers was allowed when the U.S. EPA issued its "substantially similar" definition for unleaded gasoline specifications in 1981. Under this definition, any aliphatic ether or ether mixture could be blended in unleaded gasoline as long as the total oxygen contribution from the ethers does not exceed 2.0% oxygen by weight in the gasoline. This allowed MTBE to be blended up to approximately 11 vol % in gasoline.

MTBE production grew during the 1980s as lead-based octane enhancers were being phased out of gasoline. In August 1988, U.S. EPA granted a waiver requested by Sun Refining and Marketing Co. that allowed the use of MTBE up to 15 vol %, which is approximately equal to 2.7 wt % oxygen in gasoline. This oxygen limit was extended to all aliphatic ethers in gasoline in 1990 when the U.S. EPA raised the oxygen in the "substantially similar" definition to 2.7 wt % oxygen.

In November 1990, the U.S. government again revised the Clean Air Act (CAA). One of the new requirements was for gasolines sold in ambient air quality nonattainment areas of the United States to contain a minimum amount of oxygen. For the 39 metropolitan areas that exceed the federal standard for carbon monoxide, the gasoline would have to contain at least 2.7 wt % oxygen for at least four winter months beginning November 1992. This represents nearly 30% of the U.S. gasoline market.

Starting in January 1995, the nine metropolitan areas with the most severe ozone problem must sell a "reformulated" gasoline all year long. These nine cities represent approximately 20% of the U.S. gasoline market. In addition to many other very restrictive specifications, this gasoline must also contain at least 2.0 wt % oxygen. Other metropolitan areas exceeding the federal ozone standard are also allowed to choose the program when sufficient capacity exists to make the reformulated gasoline.

Manufacture. MTBE is easily made by the selective reaction of isobutylene and methanol over an acidic ion-exchange resin catalyst, in the liquid phase and at temperatures below 100°C. To be economically competitive, MTBE's use as an octane enhancer in gasoline has been dependent on low cost isobutylene. There are a number of possible isobutylene sources for making MTBE (see BUTYLENES). During the 1980s, much of the MTBE was made with isobutylene contained in mixed butanes butylenes process streams that was produced from petrochemical olefin plants or refinery fluid catalytic crackers. MTBE plant sizes from these feedstock sources are limited by the amount of by-product isobutylene produced in these operations, and are usually in the range of one to six thousand barrels per day (40 to 250 t yr). Larger amounts of isobutylene for MTBE can be made from butanes by first isomerizing the normal butanes to isobutane, and then dehydrogenating to isobutylene. This is a much more capital intensive process, and therefore the plant sizes in the United States are usually at least 500,000 t yr to be economically competitive. Even at this size, it is still a relatively higher cost source of MTBE than that made from by-product isobutylene in ethylene plants or refineries. However, because of the growing need to meet the gasoline oxygen requirements for the U.S. Clean Air Act, many of these world-scale MTBE plants will need to be built (24).

Production. MTBE production capacity has grown steadily, usually at an annual rate of 10 to 20% per year. In 1980, world capacity was 30 thousand barrels per day (1.5 × 10⁶ t yr). By 1990, capacity was up to 180 thousand barrels per day (7 × 10⁶ t yr). Because of the requirements of the U.S. CAA, production capacity is expected to more than double from 1990 to 1995 (25). By 2000, MTBE may be the second largest organic chemical produced in the United States, second only to ethylene (26).

Economic Aspects. Prior to the 1990 revision of the CAA, MTBE sold as an octane enhancer for gasoline. Its market value has therefore been closely tied to the industry's cost of gasoline and octane. To reflect its higher-than-gasoline octane, it sold at a price higher than gasoline, usually in the range of 120 to 160% of gasoline. This price was not usually high enough to economically justify new world-scale merchant capacity based on converting

butanes and methanol to MTBE.

Because there does not appear to be sufficient MTBE capacity to satisfy the oxygen requirements for the 1990 CAA revision, there has been a general upward movement in MTBE market prices that should provide the incentive for additional capacity to be built (27).

Specifications. For fuel-grade MTBE, the typical specification of merchant product is 95 wt % minimum purity in the United States and 98% minimum in Europe. The maximums for impurities are usually 0.5 wt % for methanol, and 0.15 wt % for water. To maintain quality, fuel-grade MTBE is normally stored in fixed-roof tanks, to keep rain water out, with internal floaters, to minimize vapor losses.

Besides its use in gasoline, a high purity refined grade (99.9% min) of MTBE is produced for other industrial uses.

Uses. MTBE and related ethers are used to add octane to gasoline. It also adds oxygen to the gasoline, which allows for more efficient combustion, and therefore less carbon monoxide and unburned hydrocarbon in the exhaust emissions (11,24).

A refined grade of MTBE is used in the solvents and pharmaceutical industries. The main advantage over other ethers is its uniquely stable structural framework that contains no secondary or tertiary hydrogen atoms, which makes it very resistive to oxidation and peroxide formation. In addition, its higher autoignition temperature and narrower flammability range also make it relatively safer to use compared to other ethers (see Table 3).

The refined grade's fastest growing use is as a commercial extraction solvent and reaction medium. Other uses are as a solvent for radical-free copolymerization of maleic anhydride and an alkyl vinyl ether, and as a solvent for the polymerization of butadiene and isoprene using lithium alkyls as catalyst. Other laboratory applications include use as a solvent for Grignard reagents, and also for phase-transfer catalysts.

One other unique use of MTBE is a medical procedure for the removal of gallstones. This alternative to gallbladder surgery was developed at the Mayo Clinic, and takes advantage of MTBE's capability to quickly dissolve cholesterol. A small incision is used to inject a small amount of MTBE directly into the gallbladder, and the gallstone can then be removed in solution form (28).

TETRAHYDROFURAN

Except for the special case of the epoxides, THF represents the largest use of a heterocyclic ether. Its consumption is also much larger than that of ethyl ether. Unlike the dialkyl ethers, THF is totally miscible in water at ambient conditions. Its cyclic structure also allows it to be more reactive than the dialkyl ethers. More than half of the THF produced is used as an intermediate in making other chemicals or elastomers.

Manufacturing. Almost all the THF in the United States is currently produced by the acid-catalyzed dehydration of 1,4-butanediol [10-63-4]. Only one plant in the United States still makes THF by the hydrogenation of furfural (29). Du Pont recently claimed a new low cost process for producing THF from *n*-butane that they plan to commercialize in 1995 (30–32). The new process transport-bed oxidizes *n*-butane to crude maleic anhydride, then follows with a hydrogen reduction of aqueous maleic acid to THF (30).

Production and Specification. In 1990, approximately 105 × 10⁶ kg of THF were produced from a U.S. annual capacity of about 123 × 10⁶ kg. Consumption is expected to grow by 5 to 6% per year through 1995. THF is generally shipped in tank cars, tank trucks, or drums (29,31).

The general industry specifications are a minimum purity of 99.9 wt %. Maximum impurities are 300 ppm water and 150 ppm of the hydroperoxide of THF. Like most ethers, it is prone to forming peroxides on storage, and therefore is usually sold containing an antioxidant such as BHT.

Uses. Approximately 70% of the U.S. production is used to make poly(tetramethylene ether glycol) [25190-06-1] (PTMEG), also known as poly-THF, which is used in the production of urethane elastomers, polyurethane fibers (ether-based spandex), and copolyester–ether elastomers. PTMEG is also the fastest growing use (see POLYETHERS, TETRAHYDROFURAN). The remaining production is used as a solvent for the manufacture of poly(vinyl chloride) cements and coating, precision magnetic tape, a reaction solvent in the production of pharmaceuticals, and other miscellaneous uses.

OTHER ETHERS

***n*-Butyl Ether.** *n*-Butyl ether is prepared by dehydration of *n*-butyl alcohol by sulfuric acid or by catalytic dehydration over ferric chloride, copper sulfate, silica, or alumina at high temperatures. It is an important solvent for Grignard reagents and other reactions that require an anhydrous, inert medium. *n*-Butyl ether is also an excellent extracting agent for use with aqueous systems owing to its very low water-solubility.

Isopropyl Ether. Isopropyl ether is manufactured by the dehydration of isopropyl alcohol with sulfuric acid. It is obtained in large quantities as a by-product in the manufacture of isopropyl alcohol from propylene by the sulfuric acid process, very similar to the production of ethyl ether from ethylene. Isopropyl ether is of moderate importance as an industrial solvent, since its boiling point lies between that of ethyl ether and acetone. Isopropyl ether very readily forms hazardous peroxides and hydroperoxides, much more so than other ethers. However, this tendency can be controlled with commercial antioxidant additives. Therefore, it is also being promoted as another possible ether to be used in gasoline (33).

Tertiary-Amyl Methyl Ether. Like MTBE, TAME is produced by the simple reaction of methanol and isoamylenes (2-methyl-1-butene and 2-methyl-2-butene). It is used for gasoline blending, and also as a feedstock to make high purity isoamylenes. Its use in gasoline octane blending has not grown nearly as fast as MTBE because of less favorable economics (TAME's lower octane number). However, with the future need for gasolines that have lower vapor pressure and contain oxygen, TAME capacity is expected to grow using isoamylenes produced by refinery catalytic crackers (24).

Ethyl Tertiary-Butyl Ether. Similar to methanol in the MTBE reaction, ethanol can react with isobutylene to produce ETBE. Which alcohol is used to make the ether is highly dependent on the relative cost of the alcohols. To make ethanol more economically competitive with methanol, the federal tax credit for biomass-based ethanol used in fuel also applies to ethanol used to make ETBE in the United States (24).

Vinyl Ether. Vinyl ether is manufactured by the pyrolytic dehydrochlorination of 1,1'-dichloroethyl ether. Vinyl ether is used as a general inhalation anesthetic for procedures of short duration. Approximately 4% ethanol is added to the vinyl ether used as an anesthetic to reduce ice formation in the masks used for administration (see also VINYL POLYMERS).

2-Methoxyphenol. This ether is prepared by methylating 1,2-dihydroxybenzene. It is useful as an antioxidant for fats, oils, and vitamins, and as a polymerization inhibitor. 2-Methoxyphenol is effective as an antioxidant and inhibitor at levels of 10–100 ppm.

Butylated Hydroxyanisole. 2- and 3-*tert*-Butyl-4-methoxyphenol (butylated hydroxyanisole (BHA)) is prepared from 4-methoxyphenol and *tert*-butyl alcohol over silica or alumina at 150°C or from hydroquinone and *tert*-butyl alcohol or isobutene, using an acid catalyst and then methylating. It is widely used in all types of foods such as butter, lard, and other fats, meats, cereals, baked goods, candies, and beer as an antioxidant (see ANTIOXIDANTS; FOOD ADDITIVES). Its antioxidant properties are not lost during cooking so that flour, fats, and other BHA-stabilized ingredients may be used to produce stabilized products.

BIBLIOGRAPHY

"Ethers" in *ECT* 1st ed., Vol. 5, pp. 858–876, by M. A. Magill, Chemical Abstracts, and J. G. Parks and C. M. Beamer, Enjay Co., Inc; in *ECT* 2nd ed., Vol. 8, pp. 470–, by A. P. Lurie, Eastman Kodak Co.; in *ECT* 3rd ed., Vol. 9, pp. 381–393, by D. E. Keeley, General Electric Co.

1. S. Patai, ed., *The Chemistry of the Ether Linkage*, Wiley-Interscience, New York, 1967.
2. *Acropure High Purity MTBE Product Bulletin*, ARCO Chemical Co., Newtown Square, Pa., 1990.
3. E. W. Flick, *Industrial Solvents Handbook*, 3rd ed., Noyes Data Corp., Park Ridge, N.J., 1985.
4. T. O. Wallington and co-workers, *Int. J. Chem. Kinet.* **20**, 41–49 (1988).
5. T. O. Wallington and co-workers, *Environ. Sci. Technol.* **25**(3) 415–420 (1991).
6. E. Staude and F. Tatat in Ref. 1, Chapt. 2.
7. A. G. Davies, *Organic Peroxides*, Butterworths, London, 1961, p. 79.
8. T. W. Evans and K. R. Edlund, *Ind. Eng. Chem.* **28**(10), 1186–1188.
9. S. Scaries, Jr., and M. Tamres in Ref. 1, Chapt. 6.
10. D. L. Dalrymple, T. L. Kruger, and W. N. White in Ref. 1, Chapt. 14.

11.

W. J. Piel, *Oil Gas J.*, 40–44 (Dec. 4, 1989).

12.

C. A. Buehler and D. E. Pearson, *Survey of Organic Syntheses*, Vol. 1, Wiley-Interscience, New York, 1970, Chapt. 6.

13.

C. A. Buehler and D. E. Pearson, *Survey of Organic Syntheses*, Vol. 2, Wiley-Interscience, New York, 1977, Chapt. 6.

14.

H. L. Brockwell and co-workers, *Hydrocarbon Process.*, 133–141 (Sept. 1991).

15.

Eur. Pat. Appl. 883,008,661.3 (Sept. 9, 1988), (to Mobil Oil Corp.).

16.

Chem. Mark. Rep., (Jan. 27, 1992).

17.

MTBE/Oxygenates/ New Fuels, weekly newsletter, DeWitt & Co., Houston, Tex., Jan. 1992.

18.

M. N. Gleason, R. F. Gosselin, and H. F. Hodge, *Clinical Toxicology of Commercial Products*, The Williams and Wilkins Co., Baltimore, Md., 1957.

19.

E. G. Young and L. B. Wollner, *J. Ind. Hyg. Toxicol.* **28**, 267 (1947).

20.

A. J. Gordon and R. A. Ford, *The Chemist's Companion*, Wiley-Interscience, New York, 1972, p. 427.

21.

Ethyl Ether, *Chem. Economics Handbook*, SRI International, Menlo Park, Calif., Jan. 1991.

22.

The Pharmacopeia of the United States of America, 19th ed., USP XIX, Mack Publishing Co., Easton, Pa., 1975; USP'XX-NFXF, 1980.

23.

Chemical Safety Data Sheet SD-29, Manufacturing Chemist Association, Washington, D.C., 1972.

24.

W. J. Piel, *Fuel Reformulation* **2**(6) (1992).

25.

MTBE/Oxygenates, 1991 annual, DeWitt & Co., Houston, Tex.

26.

MTBE/Oxygenates/ New Fuels, weekly newsletter, DeWitt & Co., Houston, Tex., Aug. 8, 1991.

27.

Chem. Eng., 57 (Oct. 1991).

28.

"A Compound To Dissolve Gallstones," *New York Times*, July 23, 1988, p. 36.

29.

1,4-Butanediol, *Chem. Economics Handbook, SRI International*, Menlo Park, Calif., Sept. 1991.

30.

Chem. Mark. Rep., 3 and 14 (Sept. 3, 1990).

31.

Chem. Mark. Rep., 9 (June 24, 1991).

32.

Eur. Chem. News, 218 (Mar. 9, 1992).

33.

Oil Gas J., 39–41 (May 25, 1992).

Lawrence Karas
W. J. Piel
ARCO Chemical Company

ETHYL ALCOHOL.

See ETHANOL.

ETHYLBENZENE.

See STYRENE; XYLENES AND ETHYLBENZENE.

ETHYLENE

Ethylene [74-85-1] (ethene), H₂C=CH₂, is the largest volume building block for many petrochemicals. This olefin is used to produce many end products such as plastics, resins, fibers, etc. Ethylene is produced mainly from petroleum-based feedstocks by thermal cracking, although alternative methods are also gaining importance.

Physical Properties

Ethylene is the lightest olefin. It is a colorless, flammable gas with a slightly sweet odor. Physical and thermodynamic properties are given in many references (1–7), and are briefly summarized in Tables 1–3.

Table 1. Physical Properties of Ethylene

Property	Value
mol wt	28.0536
triple point	
temperature, °C	169.164
pressure, kPa ^a	0.12252
latent heat of fusion, kJ mol ^b	3.353
normal freezing point	
temperature, °C	169.15
latent latent heat of fusion, kJ mol ^b	3.353
normal boiling point	
temperature °C	103.71
latent heat of vaporization, kJ mol ^b	13.548
density of liquid	
mol L	20.27
<i>d</i> ₄ ¹⁰⁴	0.566
specific heat of liquid, J (mol·K) ^b	67.4
viscosity of the liquid, mPa·s (cP)	0.161
surface tension of the liquid, mN m (dyn cm)	16.4
specific heat of ideal gas at 25°C, J (mol·K) ^b	42.84
critical point	
temperature, °C	9.194
pressure, kPa ^a	5040.8
density, mol L	7.635
compressibility factor	0.2812
gross heat of combustion of gas at 25°C, MJ mol ^b	1.411
limits of flammability at atmospheric pressure and 25°C	
lower limit in air, mol %	2.7
upper limit in air, mol %	36.0
autoignition temperature in air at atmospheric pressure, °C	490.0
Pitzer's accentric factor	0.278
standard enthalpy of formation at 25°C, kJ mol ^b	52.3
standard Gibbs energy of formation at 25°C for ideal gas at atmospheric pressure, kJ mol ^b	68.26
solubility in water at 0°C and 101 kPa ^a , mL mL H ₂ O	0.226
speed of sound at 0°C and 409.681 kPa ^a , m s	224.979
standard entropy of formation, J (mol·K) ^b	219.28
standard heat capacity, J (mol·K) ^b	42.86

^a To convert kPa to mm Hg, multiply by 7.5.

^b To convert J to cal, divide by 4.184.

Table 2. Thermodynamic and Transport Properties of Gaseous Ethylene

Temperature, °C	Specific volume, L kg	Viscosity, mPa·s (cP)	Thermal cond, $\frac{MW}{(m\cdot K)}$	Heat capacity, C_p , J (mol·K) ^a
<i>Pressure 100 kPa^b</i>				
0°C	803.50	0.00941		40.90
25°C	878.63	0.01027	20.3	43.11
50°C	953.52	0.01108	23.3	45.48
75°C	1028.20	0.01180	26.7	47.95
100°C	1102.90	0.01260	30.2	50.43
<i>Pressure 5000 kPa^b</i>				
25°C	11.10	0.01404	31.5	92.8
50°C	14.33	0.01310	30.3	67.2
75°C	16.82	0.01314	31.7	61.5
100°C	19.02	0.01395	34.9	60.0

^a To convert J to cal, divide by 4.184.
^b To convert kPa to psi, multiply by 0.145.

Table 3. Thermodynamic and Transport Properties of Liquid Ethylene

Temp-erature, °C	Vapor pressure, kPa ^a	Density, mol L	Viscosity, mPa·s (cP)	Thermal conductivity, W (m·K)	Heat capacities	
					C_p , J (mol·K) ^b	C_v , J (mol·K) ^b
165°C	0.249	23.16	0.600	0.252		44.75
150°C	2.040	22.47	0.387	0.237	70.66	45.55
125°C	23.76	21.29	0.229	0.214	66.76	39.54
100°C	126.0	20.05	0.160	0.191	67.44	37.55
75°C	422.4	18.70		0.166	70.68	37.13
50°C	1063	17.15		0.140	77.15	37.39
25°C	2219	15.20		0.113	92.57	38.44
0°C	4100	12.20		0.080	185.00	41.82

^a To convert kPa to mm Hg, multiply by 7.5.
^bTo convert J to cal, divide by 4.184.

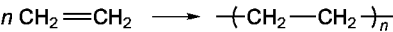
Chemical Properties

Structure. Ethylene is a planar molecule with a carbon–carbon bond distance of 0.134 nm, which is shorter than the C–C bond length of 0.153 nm found in ethane. The C–H bond distance is 0.110 nm, and the bond angles are <HCH = 117.2° and <HCC = 121.4° consistent with the *sp*² hybridized state.

Reactivity. Ethylene is a very reactive intermediate, and hence is involved in many chemical reactions. The ethylene double bond reacts readily to form saturated hydrocarbons, their derivatives, or polymers. Electrons in the π -bond are less tightly held and more easily polarized than electrons in a sigma bond. The carbon-carbon double-bond energy is 611 kJ mol (146 kcal mol), which is less than twice the C–C bond dissociation energy of 368 kJ mol (88 kcal mol) found in ethane. The C–H bond dissociation energy is 452 kJ mol (108 kcal mol), and approximate acidity as measured by *K_a* is 10^{–45}. Therefore, ethylene reacts with electrophilic reagents like strong acids (H⁺), halogens, and oxidizing agents, but not with nucleophilic reagents such as Grignard reagents and bases. Some of the reactions have commercial significance and others have only academic interest. For fundamental mechanisms of these reactions see References 8 and 9.

The principal reactions with commercial significance include polymerization, oxidation, and addition including halogenation, alkylation, oligomerization, hydration, and hydroformylation.

Polymerization. Very high purity ethylene (>99.9% plus) is polymerized under specific conditions of temperature and pressure in the presence of an initiator or catalyst.



This is an exothermic reaction, and both homogeneous (radical or cationic) and heterogeneous (solid catalyst) initiators are used. The products range in molecular weight from below 1000 to a few million (see OLEFIN POLYMERS). Reaction mechanisms and reactor designs have been extensively discussed (10–12).

There are four types of reaction systems for the production of polyethylene of commercial importance:

- (1) High pressure (60–350 MPa) free-radical polymerization using oxygen, peroxide, or other strong oxidizers as initiators at temperatures of up to 350°C to produce low density polyethylene (LDPE), a highly branched polymer, with densities from 0.91 to 0.94 g cm³.
- (2) Low pressure (0.1 to 20 MPa) and temperatures of 50 to 300°C using heterogeneous catalysts such as molybdenum oxide or chromium oxide supported on inorganic carriers to produce high density polyethylene (HDPE), which is more linear in nature, with densities of 0.94 to 0.97 g cm³.
- (3) Low pressure polymerization via ionic catalysts, using Ziegler catalysts (aluminum alkyls and titanium halides).
- (4) Low pressure polymerization with Ziegler catalysts supported on inorganic carriers.

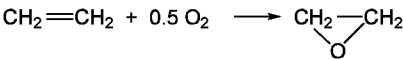
A more recent development in ethylene polymerization is the simplified low pressure LDPE process. The pressure range is 0.7–2.1 MPa with temperatures less than 100°C. The reaction takes place in the gas phase instead of liquid phase as in the conventional LDPE technology. These new technologies demand ultra high purity ethylene.

In order to improve the physical properties of HDPE and LDPE, copolymers of ethylene and small amounts of other monomers such as higher olefins, ethyl acrylate, maleic anhydride, vinyl acetate, or acrylic acid are added to the polyethylene. For example, linear low density polyethylene (LLDPE), although linear, has a significant number of branches introduced by using comonomers such as 1-butene or 1-octene. The linearity provides strength, whereas branching provides toughness.

Ethylene–propylene elastomers exhibit a high resistance to oxygen, ozone, and heat, and are used in motor parts because of their excellent low

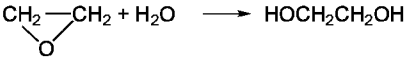
temperature flexibility. Two types of elastomers that are manufactured are copolymers (EPR) formed by polymerizing ethylene and propylene, and terpolymers (EPDM) formed by polymerizing ethylene, propylene, and a small amount of a termonomer like a diene or triene, eg, 1,4-hexadiene (see ELASTOMERS, SYNTHETIC ETHYLENE PROPYLENE DIENE RUBBER).

Oxidation. Ethylene oxide (qv) is produced by oxidizing ethylene.

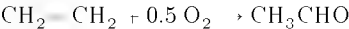


The reaction is carried out over a supported metallic silver catalyst at 250–300°C and 1–2 MPa (10–20 bar). A few parts per million (ppm) of 1,2-dichloroethane are added to the ethylene to inhibit further oxidation to carbon dioxide and water. This results in chlorine generation, which deactivates the surface of the catalyst. Chem Systems of the United States has developed a process that produces ethylene glycol monoacetate as an intermediate, which on thermal decomposition yields ethylene oxide [75-21-8].

About 60% of the ethylene oxide produced is converted to ethylene glycol by reaction of ethylene oxide in the presence of excess water and an acidic catalyst at 50–70°C. This is followed by hydrolysis at relatively high temperatures (140–230°C) and 2–4 MPa (20–40 bar) (see GLYCOLS, ETHYLENE GLYCOL). When the water concentration is lowered, poly(ethylene glycol) is obtained.

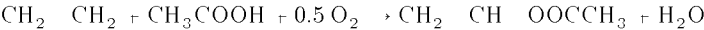


Acetaldehyde [75-07-0] can be obtained by the Wacker process, in which a homogeneous CuCl₂–PdCl₂ system is used for the oxidation.



The reaction is carried out in a bubble column at 120–130°C and 0.3 MPa (3 bar). Palladium chloride is reduced to palladium during the reaction, and then is reoxidized by cupric chloride. Oxygen converts the reduced cuprous chloride to cupric chloride.

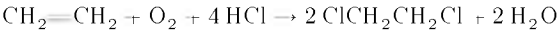
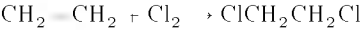
Vinyl acetate [108-05-4] is obtained by vapor-phase oxidation of ethylene with acetic acid. Acetic acid is obtained by oxidation of acetaldehyde.



This process employs a palladium on carbon, alumina, or silica–alumina catalyst at 175–200°C and 0.4 to 1.0 MPa (58–145 psi).

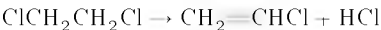
Addition. Addition reactions of ethylene have considerable importance and lead to the production of ethylene dichloride, ethylene dibromide, and ethyl chloride by halogenation–hydrohalogenation; ethylbenzene, ethyltoluene, and aluminum alkyls by alkylation; α-olefins by oligomerization; ethanol by hydration; and propionaldehyde by hydroformylation.

Halogenation–Hydrohalogenation. The most important intermediate is ethylene dichloride [107-06-2] (EDC) which is produced from ethylene either by direct chlorination or by oxychlorination. Direct chlorination is carried out in the liquid or vapor phase over catalysts of iron, aluminum, copper, or antimony chlorides, and at conditions of 60°C. Oxychlorination is carried out in a fixed or fluidized bed at 220°C with a suitable solid chloride catalyst.



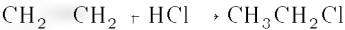
By similar methods, ethylene dibromide [106-93-4] can also be obtained. This compound is used in gasoline as an antiknocking additive.

The leading derivative of ethylene dichloride is vinyl chloride [75-01-4] monomer (VCM), which is subsequently used to produce poly(vinyl chloride) and chlorinated hydrocarbons. Vinyl chloride is obtained by dehydrochlorination of ethylene dichloride in the gas phase (500–600°C and 2.5–3.5 MPa).



Trichloro- and tetrachloroethylene are important organic solvents. These are produced by further chlorination of 1,2-dichloroethylene in the gas phase with simultaneous dehydrochlorination in the presence of a suitable chloride catalyst (see CHLOROCARBONS AND CHLOROHYDROCARBONS).

Ethyl chloride [25-00-3] is obtained by reaction of ethylene with hydrogen chloride in the presence of AlCl₃ or FeCl₃ at 300–500 kPa (3–5 bar) and 30–90°C (liquid phase), or at 130–250°C (vapor phase).



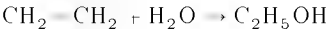
Alkylation. Ethylbenzene [100-41-4], the precursor of styrene, is produced from benzene and ethylene. The ethylation of benzene is conducted either in the liquid phase in the presence of a Friedel-Crafts catalyst (AlCl₃, BF₃, FeCl₃) or in the vapor phase with a suitable catalyst. The Monsanto–Lummus process uses an aluminum chloride catalyst that yields more than 99% ethylbenzene (13). More recently, Lummus and Union Oil commercialized a zeolite catalyst process for liquid-phase alkylation (14). Badger and Mobil also have a vapor-phase alkylation process using zeolite catalysts (15). Almost all ethylbenzene produced is used for the manufacture of styrene [100-42-5], which is obtained by dehydrogenation in the presence of a suitable catalyst at 550–640°C and relatively low pressure, <0.1 MPa (<1 atm).

Toluene also reacts with ethylene to produce *p*-ethyltoluene [622-96-8] or *p*-methylethylbenzene, which can be dehydrogenated to give *p*-methylstyrene. The polymer (PMS), has a high glass-transition point and better flow properties, and has gained significant commercial importance in recent years.

Alkylation of aluminum with ethylene yields products that find application as initiators and starter compounds in the production of α-olefins and linear primary alcohols, as polymerization catalysts, and in the synthesis of some monomers like 1,4-hexadiene. Triethylaluminum [97-93-8], Al(C₂H₅)₃, is the most important of the ethylene-derived aluminum alkyls.

In the production of α-olefins, ethylene reacts with an aluminum alkyl at relatively low temperature to produce a higher alkylaluminum. This is then subjected to a displacement reaction with ethylene at high temperatures to yield a mixture of α-olefins and triethylaluminum. In an alternative process, both reactions are combined at high temperatures and pressures where triethylaluminum functions as a catalyst in the polymerization process.

Hydration. Ethanol [64-17-5] is manufactured from ethylene by direct catalytic hydration over a H₃PO₄–SiO₂ catalyst at process conditions of 300°C and 7.0 MPa (1015 psi). Diethyl ether is also formed as a by-product.



In another process, ethylene is absorbed in 90–98% sulfuric acid at 50–85°C and 1.0–1.4 MPa (145–200 psi) to give a mixture of ethyl sulfates. These can be hydrolyzed to ethanol and dilute sulfuric acid.

Hydroformylation. In hydroformylation, the oxo reaction, ethylene reacts with synthesis gas (CO + H₂) over a cobalt catalyst at 60–200°C and 4–35 MPa (39–345 atm) to form propionaldehyde [123-38-6] (see OXO PROCESS). This reaction is usually carried out in an organic liquid phase where inert diluents are used as liquids (16).

Other Reactions. Comprehensive discussions of the following reactions, which are primarily of academic interest, are provided in various references (10,17). Ethylene may be hydrogenated to ethane under a variety of conditions. For example, hydrogenation is feasible for systems utilizing finely

divided platinum or palladium at room temperature and atmospheric pressure, or utilizing nickel at 150–300°C and elevated pressure. These reactions are mainly used in research work.

Acyl halides may also be added to ethylene in the presence of aluminum chloride to form halogenated ketones. At low temperatures, ethylene reacts with halogens to yield dihaloethanes. At high temperatures, trichloroethylene and perchloroethylene are formed. The most profitable route for chloroethylene is via ethylene dichloride (see CHLOROCARBONS AND CHLOROHYDROCARBONS).

Biological Properties

Ethylene is slightly more potent as an anesthetic than nitrous oxide, and the smell of ethylene causes choking. Diffusion through the alveolar membrane is sufficiently rapid for equilibrium to be established between the alveolar and the pulmonary capillary blood with a single exposure. Ethylene is held both in cells and in plasma in simple physical solution. The lipid stroma of the red blood cells absorb ethylene, but it does not combine with hemoglobin. The concentration in the blood is 1.4 mg mL when ethylene is used by itself for anesthesia. However, in the 1990s it is not used as an anesthetic agent. Ethylene is eliminated from the body unchanged, primarily by the lungs, and most elimination is complete within three minutes of administration.

The controlled ripening of various fruits and vegetables by ethylene is of considerable importance. Ethylene was identified as one of the volatiles emitted by ripening fruits as early as the 1930s, and its biological use had been mentioned as early as 1901. A few ppm of ethylene (<10 ppm) is often used for ripening bananas and other fruits. However, the ethylene concentration varies with the type of fruit and environmental conditions. Ethylene can also have adverse effects on plants. It causes the bleaching of green tissue, gives rise to foliar abscission, suppresses certain types of dormancy, and promotes cellular swelling. For further information on this subject, consult Reference 10.

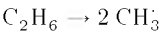
Manufacture by Thermal Cracking

Although ethylene is produced by various methods as follows, only a few are commercially proven: thermal cracking of hydrocarbons, catalytic pyrolysis, membrane dehydrogenation of ethane, oxydehydrogenation of ethane, oxidative coupling of methane, methanol to ethylene, dehydration of ethanol, ethylene from coal, disproportionation of propylene, and ethylene as a by-product.

Thermal cracking of hydrocarbons is the principal route for the industrial production of ethylene. The chemistry and engineering of thermal cracking has been reviewed (17–19). In thermal cracking, valuable by-products including propylene, butadiene, and benzene are also produced. Commercially less valuable methane and fuel oil are also produced in significant proportions. An important parameter in the design of commercial thermal cracking furnaces is the selectivity to produce the desired products.

Mechanism. The thermal cracking of hydrocarbons proceeds via a free-radical mechanism (20). Since that discovery, many reaction schemes have been proposed for various hydrocarbon feeds (21–24). Since radicals are neutral species with a short life, their concentrations under reaction conditions are extremely small. Therefore, the integration of continuity equations involving radical and molecular species requires special integration algorithms (25). An approximate method known as pseudo steady-state approximation has been used in chemical kinetics for many years (26,27). The errors associated with various approximations in predicting the product distribution have been given (28).

The free-radical mechanism involves initiation, propagation, and termination steps. During *initiation* two radicals are produced for every paraffin molecule.



Only a small fraction of reactant is involved in this step. When naphthenes are involved, diradicals are produced. For aromatics with side chains, H• radicals are produced.

In *propagation* many types of reactions are involved including H abstraction, addition, radical decomposition, and radical isomerization.

In H abstraction, a hydrogen radical reacts with a molecule (primarily a paraffin) and produces a hydrogen molecule and a radical. In the same way, a methyl radical reacts to produce a radical and methane. Similar reactions with other radicals (ethyl and propyl) can also occur. In addition, some radicals like H•, CH₃•, etc, are added to olefins to form heavier radicals.

Radical decomposition is one of the most important types of reactions. In this case, a larger radical decomposes to an olefin and a smaller radical. Radicals usually decompose at the beta position of the radical center where the C–C bond is the weakest. In the case of naphthenes and aromatics this may not be the case, and C–H bond may be the weakest. Radical isomerization frequently occurs for large radicals, and explains to a large extent the observed product distribution.

Radical termination is the reverse of initiation.

Conversion. The terms severity or conversion are used to measure the extent of cracking. Conversion can easily be measured for a single component (C) feed and is defined as follows, where the quantities are measured in weight units.

conversion = $\frac{C_{in} - C_{out}}{C_{in}}$

When a mixture is cracked, one or more components in the feed may also be formed as products. For example, in the cocracking of ethane and propane, ethane is formed as a product of propane cracking and propane is formed as a product of ethane cracking. Therefore, the "out" term in the above equation contains the contribution or formation from other feed components and hence does not represent true conversion. For simple mixtures, the product formation can be accounted for, and approximate true conversions can be calculated (29). For liquid feeds like naphtha, it is impractical if not impossible to calculate the true conversion. Based on measured feed components, one can calculate a weighted average conversion (X) (30):

X = ΣW_i X_i

where X_i is the conversion for the *i*th feed component, and W_i is the weighting factor (weight or mole fraction, usually).

As a practical method, designers have employed other methods such as *n*-pentane conversion as a key component, kinetic severity factor (31), or molecular collision parameter (32) to represent severity. Alternatively, molecular weight of the complete product distribution has been used to define conversion (X) for liquid feeds.

X = $\frac{\left[\left(\frac{MW_f}{MW_e} \right) - 1 \right]}{\left[\left(\frac{MW_f}{24.5} \right) - 1 \right]}$

In the equation, MW_f and MW_e are the molecular weight of the feed and of the dry (steam-free) effluent, respectively. Instead of molecular weight, hydrogen content in the C₅-plus product is also used.

Instead of conversion, some producers prefer to use other identifications of severity, including coil outlet temperature, propylene to methane ratio, propylene to ethylene ratio, or cracking severity index (33). Of course, all these definitions are somewhat dependent on feed properties, and most also depend on the operating conditions.

When simple liquids like naphtha are cracked, it may be possible to determine the feed components by gas chromatography combined with mass spectrometry (gc ms) (30). However, when gas oil is cracked, complete analysis of the feed may not be possible. Therefore, some simple definitions are used to characterize the feed. When available, paraffins, olefins, naphthenes, and aromatics (PONA) content serves as a key property. When PONA is not available, the Bureau of Mines Correlation Index (BMCI) is used. Other properties like specific gravity, ASTM distillation, viscosity, refractive index. Conradson Carbon, and Bromine Number are also used to characterize the feed. In recent years even nuclear magnetic resonance spectroscopy has been

used to characterize heavy feedstocks.

Industrial Furnaces. Thermal cracking of hydrocarbons is accomplished in tubular reactors commonly known as cracking furnaces, crackers, cracking heaters, etc. Several engineering contractors including ABB Lummus Crest, Stone and Webster, M. W. Kellogg, Linde, C. F. Braun, and KTI offer cracking furnace technology. Usually two cracking furnaces share a common stack, and the height of the heater may vary from 30 to 50 m. Before the 1960s, the cracking tubes were arranged in horizontal rows in a radiant chamber leading to low ethylene capacity (<20,000 t yr). Modern designs use tubes arranged in vertical rows, providing superior mechanical performance and higher capacity. The capacity of a single furnace is well over 100,000 t yr. A typical sketch of such a furnace is shown in Figure 1.

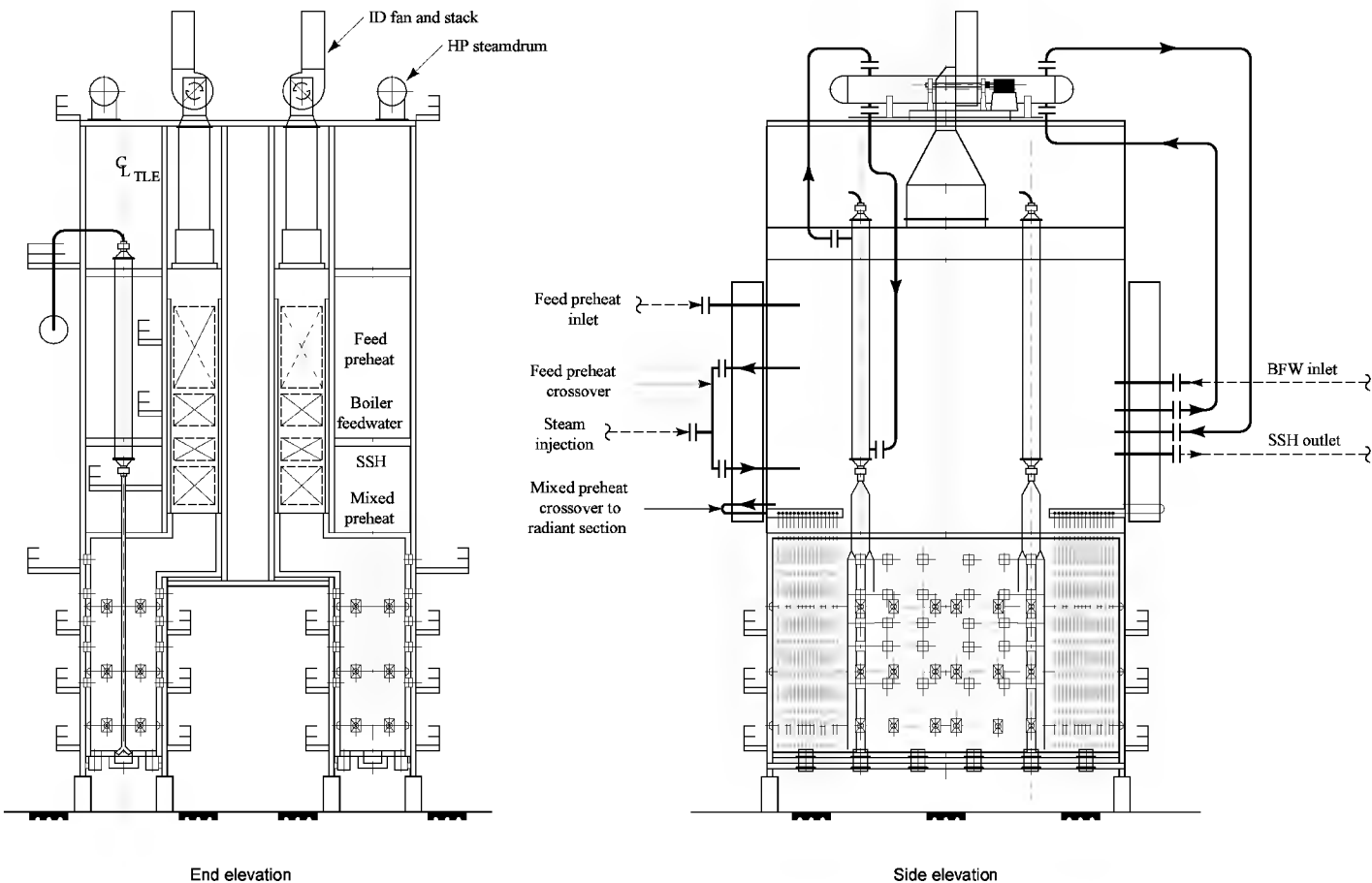


Fig. 1. Typical heater configuration. BFW = boiler feed water; SSH = super high pressure steam; HP = high pressure; and ID = induced draft..

Reaction. The reaction proceeds in the pyrolysis coils of the radiant section of the furnace. Since coke is also formed during pyrolysis, steam is added as a diluent to the feed. The steam minimizes the side reaction forming coke, and improves selectivity to produce the desired olefins by lowering hydrocarbon partial pressure. The temperature of the hydrocarbon and steam mixture entering the radiant chamber (known as the crossover temperature) is 500 to 700°C. Lower temperatures are used for heavy feeds like atmospheric gas oil (AGO) and vacuum gas oils (VGO), and higher temperatures are used for light gases like ethane and propane. Some cracking can start as low as 400°C. However, for light gases incipient conversion is quite low. Depending on the residence time and required feed severity, the coil outlet temperature is typically maintained between 775 and 950°C.

The combination of low residence time and low partial pressure produces high selectivity to olefins at a constant feed conversion. In the 1960s, the residence time was 0.5 to 0.8 seconds, whereas in the late 1980s, residence time was typically 0.1 to 0.15 seconds. Typical pyrolysis heater characteristics are given in Table 4. Temperature, pressure, conversion, and residence time profiles across the reactor for naphtha cracking are illustrated in Figure 2.

Table 4. Pyrolysis Heater Characteristics

Single heater characteristics	Range
number of coils	2–176
coil length, m	9–80
inside coil diameter, mm	30–200
process gas outlet temperature, °C	750–950
clean coil metal temperature, °C	900–1,080
max metal temperature, °C	1,040–1,150
average heat absorption, kW m ² ext. area	50–110
bulk residence time, s	0.1–0.6
coil outlet pressure, kPa ^a	150–275
clean coil pressure drop, kPa ^a	10–200
ethylene capacity, t yr	20,000–100,000

^a To convert kPa to bar, divide by 100.

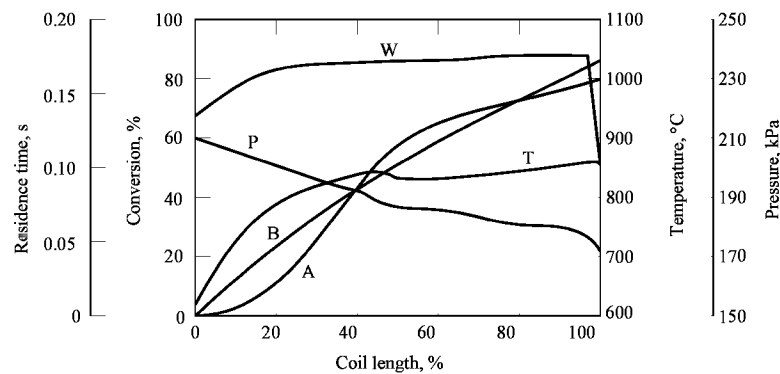


Fig. 2. Profiles of conversion, A, having pressure P and gas temperature T along the reactor length. Residence time, B, where W is maximum wall temperature. To convert kPa to bar, divide by 100.

5

Cracking reactions are endothermic, 1.6–2.8 MJ/kg (700–1200 BTU/lb) of hydrocarbon converted, with heat supplied by firing fuel gas and/or fuel oil in side-wall or floor burners. Side-wall burners usually give uniform heat distribution, but the capacity of each burner is limited (0.1–1 MW) and hence 40 to 200 burners are required in a single furnace. With modern floor burners, also called hearth burners, uniform heat flux distribution can be obtained for coils as high as 10 m, and these are extensively used in newer designs. The capacity of these burners varies considerably (1–10 MW), and hence only a few burners are required. The selection of burners depends on the type of fuel (gas and/or liquid), source of combustion air (ambient, preheated, or gas turbine exhaust), and required NO_x levels.

The reaction mixture exiting the furnace is quickly cooled in quench coolers also called transferline exchangers (TLE). In earlier designs, direct quenching with oil or water was used for most liquid feeds. Almost all modern designs employ indirect quenching, by which valuable high pressure steam is generated. Direct quenching is only used for very heavy feeds.

Single-stage or two-stage cooling is used to achieve the desired degree of cooling. These alternatives achieve the same thermal efficiency, but offer different approaches to low pressure drop operation and mechanical design.

In the first stage, the process gas is cooled in a double-pipe exchanger or in a shell and tube exchanger. In the second stage, a shell and tube exchanger is used to generate additional steam and sometimes to preheat the feed and dilution steam. The outlet gas temperature from the transferline exchanger varies from 350 to 650°C depending on the feedstock and the design. If the reaction mixture is not cooled quickly, olefin selectivity is reduced because of the many side reactions taking place in this zone. After the transferline exchanger, further cooling is achieved by spraying the furnace effluent with quench oil. The oil and water phases are then separated from the gas phase. The cooled furnace effluent then proceeds to the recovery section for further separation.

Efficiency. Since only 35 to 50% of fired duty is absorbed in the radiant section, the flue gas leaving the radiant chamber contains considerable energy that can be extracted efficiently in the convection section of the furnace. In the convection section, the feed is preheated along with dilution steam to the desired crossover temperature. Residual heat is recovered by generating steam. The overall thermal efficiency of modern furnaces exceeds 93%, and a value of 95% is not uncommon.

The convection section is a series of cross-flow exchangers with flue gas on one side and process fluids on the other (tube) side. Since mainly gas-to-gas heat transfer is involved, fin tubes are employed to improve the heat-transfer rate where practical. The metallurgy of the tubes varies from carbon steel to high temperature alloy depending on the service. When high overall efficiency is desired, condensation of acidic flue gases must be taken into account in the selection of materials. Fouling of heat-transfer surfaces both inside and outside is unavoidable. Outside fouling is cleaned by steam lancing, and inside fouling is usually handled by burning.

In new plant designs, cogeneration of electricity and steam has become economically attractive. Depending on the plant capacity, gas turbines (15–50 MW) are used to generate electric power. These turbines usually burn fuel gas with 200%+ excess air. Therefore, the exhaust gas is not only hot, but also rich in oxygen. Instead of directly generating steam from the exhaust gas by using waste heat boilers, the gas is fed to the cracking heaters as a source of combustion oxygen. Typically, the exhaust gas temperature is 400 to 590°C with an oxygen content of 14 mol% or more. Typical energy savings of 10 to 30% are reported (34,35). Using hot combustion air requires special ducting, and hence the investment cost of the heater is slightly higher.

Instead of gas turbine exhaust, air preheat has been used in some plants to reduce fuel consumption. Flue gas leaving the furnace stack passes through an air preheater, and the preheated air is supplied to the burners. By using mostly hearth burners, the duct work and the investment cost can be minimized with air preheat and gas turbine exhaust. It is also possible with 100% wall-fired furnaces, and has been proven in commercial operation (34).

Environmental. Stringent environmental laws require that nitrogen oxides (NO_x) and sulfur oxides emission from furnaces be drastically reduced. In many parts of the world, regulations require that NO_x be reduced to 70 vol ppm or lower on a wet basis. Conventional burners usually produce 100 to 120 vol ppm of NO_x. Many vendors (McGill, John Zink, and North American) are supplying low NO_x burners.

Since NO_x production depends on the flame temperature and quantity of excess air, achieving required limits may not be possible through burner design alone. Therefore, many new designs incorporate DENOX units that employ catalytic methods to reduce the NO_x limit. Platinum-containing monolithic catalysts are used (36). Each catalyst performs optimally for a specific temperature range, and most of them work properly around 400°C.

Product Distribution. In addition to ethylene, many by-products are also formed. Typical product distributions for various feeds from a typical short residence time furnace are shown in Table 5. The product distribution is strongly influenced by residence time, hydrocarbon partial pressure, steam-to-oil ratio, and coil outlet pressure.

Table 5. Product Distribution Obtained in a Short Residence Time Coil at 172 kPa^a

Property	C ₂ H ₄	C ₃ H ₈	<i>n</i> -C ₄ H ₁₀	Light naphtha	Light AGO	HVGO ^b
sp gr				0.662	0.8191	0.852
boiling range, °C				35–150	185–335	360–540
mol wt	30.0	44.0	58.0	81.0	205.0	425.0
feed H ₂ , wt %	20.10	18.29	17.34	16.00	13.93	14.20
steam HC, wt wt	0.30	0.30	0.40	0.50	0.75	0.75
severity conv, %	65	95	96	c	c	c
yields, wt %						
H ₂	3.93	1.56	1.17	1.00	0.63	0.65
CH ₄	3.82	25.30	21.70	18.00	11.20	12.60
C ₂ H ₂	0.43	0.64	0.78	0.95	0.47	0.33

C ₂ H ₄	53.00	39.04	39.20	34.30	26.50	29.00
C ₂ H	35.00	3.94	3.02	3.80	3.40	3.70
C ₃ H ₄	0.06	0.53	1.15	1.02	0.80	0.95
C ₃ H	0.89	11.34	15.34	14.10	13.40	13.10
C ₃ H ₈	0.17	5.00	0.16	0.35	0.25	0.24
C ₄ H	1.19	4.50	4.08	4.45	5.00	5.00
C ₄ H ₈	0.18	0.80	1.69	3.70	3.70	3.40
C ₄ H ₁₀	0.22	0.09	4.00	0.20	0.10	0.07
C ₅	0.27	1.61	1.38	2.10	2.75	1.90
C –C ₈ ^d	0.39	0.31	1.45	0.80	1.20	1.40
benzene	0.37	2.74	2.48	6.40	6.90	7.30
toluene	0.08	0.67	0.52	2.30	3.20	3.65
xylene + ethylbenzene	0.00	0.09	0.20	0.21	1.30	1.10
styrene	0.00	0.51	0.23	0.75	0.79	0.65
C ₉ –205°C	0.00	0.93	0.87	1.40	2.96	2.90
fuel oil	0.00	0.40	0.58	4.17	15.45	12.06
Total	100.00	100.00	100.00	100.00	100.00	100.00

- ^a Absolute 1.72 bar.
- ^b Hydrocracked vacuum gas oil.
- ^c Maximized ethylene product.
- ^d Nonaromatic.

Table 6 shows the effect of varying coil outlet pressure and steam-to-oil ratio for a typical naphtha feed on the product distribution. Although in these tables, the severity is defined as maximum, in a realistic sense they are not maximum. It is theoretically possible that one can further increase the severity and thus increase the ethylene yield. Based on experience, however, increasing the severity above these practical values produces significantly more fuel oil and methane with a severe reduction in propylene yield. The run length of the heater is also significantly reduced. Therefore, this is an arbitrary maximum, and if economic conditions justify, one can operate the commercial coils above the so-called maximum severity. However, after a certain severity level, the ethylene yield drops further, and it is not advisable to operate near or beyond this point because of extremely severe coking.

Table 6. Product Distribution, wt %, as a Function of Severity and Selectivity for a Naphtha Feed^{a, b, c}

Product	Max C ₃ H	Max olefin	Max C ₂ H ₄	Max C ₂ H ₄ ^d	Max C ₂ H ₄ ^e
H ₂	0.75	0.86	0.91	0.91	0.91
CH ₄	12.60	14.65	15.70	15.95	15.30
C ₂ H ₂	0.43	0.66	0.78	0.66	0.95
C ₂ H ₄	25.50	29.00	30.80	29.95	32.20
C ₂ H	4.30	4.10	3.30	3.62	2.80
C ₃ H ₄	0.56	0.80	1.00	0.91	1.15
C ₃ H	17.00	15.90	14.00	13.70	14.40
C ₃ H ₈	0.45	0.34	0.28	0.32	0.22
C ₄ H	4.50	4.75	4.70	4.55	4.90
C ₄ H ₈	6.50	4.15	3.80	3.80	3.81
C ₄ H ₁₀	0.80	0.30	0.20	0.20	0.20
C ₅	4.95	3.20	2.93	2.78	3.10
C –C ₈ non-aromatic	6.40	2.60	1.80	1.55	2.20
benzene	4.00	6.00	6.70	7.15	5.95
toluene	3.80	3.93	4.00	4.10	3.90
xylene + ethylbenzene	2.20	1.57	1.30	1.34	1.24
styrene	0.65	0.78	0.82	0.87	0.75
C ₉ –205°C	2.16	1.90	1.82	1.89	1.72
fuel oil	2.45	4.51	5.16	5.75	4.30
Total	100.00	100.00	100.00	100.00	100.00

- ^a Specific gravity 0.7260.
- ^b Coil outlet pressure (COP) 172 kPa unless otherwise noted.
- ^c Stream HC, wt wt = 0.5 unless otherwise noted.
- ^d COP 207 kPa.
- ^e Stream HC, wt wt = 0.75.

Kinetic Models Used for Designs. Numerous free-radical reactions occur during cracking; therefore, many simplified models have been used. For example, the reaction order for overall feed decomposition based on simple reactions for alkanes has been generalized (37). Many researchers have correlated the overall decomposition as an *n*th order reaction, with most paraffins following the first order and most olefins following a higher order. In general, isoparaffin rate constants are lower than normal paraffin rate constants. The rate constants are somewhat dependent on conversion due to inhibition effects; that is, the rate constant often decreases with increasing conversion, and the order of conversion is not affected. This has been explained by considering the formation of allyl radicals (38). To predict the product distribution, yields are often correlated as a function of conversion or other severity parameters (39). Instead of radical reactions, models based on molecular reactions have been proposed for the cracking of simple alkanes and liquid feeds like naphtha and gas oil (40–42). However, the validity of these models is limited, and cannot be extrapolated outside the range with confidence. With sophisticated algorithms and high speed computers available, this molecular reaction approach is not recommended. With the introduction of Gear's algorithm (25) for integration of stiff differential equations, the complete set of continuity equations describing the evolution of radical and molecular species can be solved even with a personal computer. Many models incorporating radical reactions have been published.

For the cracking of light hydrocarbons at low temperatures, one study (22) used more than 300 reactions, and another (21), 133 reactions. The estimation of frequency factors and activation energies of these elementary reactions from other thermodynamic data has been outlined (43). Properties of various radicals are available (2,44–47). A more detailed model for the cracking of hydrocarbons has been published (24). Unfortunately, these models assume pseudo steady state for the radicals, at least for a small portion of the furnace if not over the complete furnace. This integration is inferior to Gear's algorithm which does not assume steady state. For example, calculations generated for propane cracking in a non-isothermal pilot reactor with Gear's algorithm, show clearly that many radicals do not attain steady state (21).

Run Length. Coke is produced as a side product that deposits on the radiant tube walls. This limits the heat transfer to the tubes, and increases the pressure drop across the coil. The coke deposition not only limits the heat transfer, but also reduces the olefin selectivity. Periodically, the heater has to be shut down and cleaned. Typical run lengths are 40 to 100 days between decokings. Prediction of run length of a commercial furnace is still an art, and various mechanisms are postulated in the literature. Coke also deposits in transferline exchangers. Mechanisms for coking in radiant coils and transferline exchangers appears to be different for different feeds.

Over 25 years ago the coking factor of the radiant coil was empirically correlated to operating conditions (48). It has been assumed that the mass transfer of coke precursors from the bulk of the gas to the walls was controlling the rate of deposition (39). Kinetic models (24,49,50) were developed based on the chemical reaction at the wall as a controlling step. Bench-scale data (51–53) appear to indicate that a chemical reaction controls. However, flow regimes of bench-scale reactors are so different from the commercial furnaces that scale-up of bench-scale results cannot be confidently applied to commercial furnaces. For example, Figure 3 shows the coke deposited on a controlled cylindrical specimen in a continuous stirred tank reactor (CSTR) and the rate of coke deposition. The deposition rate decreases with time and attains a pseudo steady value. Though this is achieved in a matter of minutes in bench-scale reactors, it takes a few days in a commercial furnace.

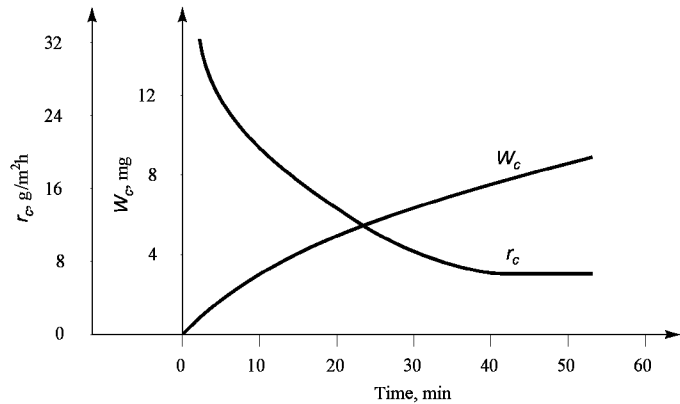


Fig. 3. Weight of coke formed (W_c) and coking rate (r_c) in ethane cracking as a function of time (51).

Transferline exchanger (TLE) coking is different from coking of the radiant coil. Condensation of coke precursors contained in the heavier fractions of the cracked gas has been claimed to account for TLE fouling (54). Mass transfer of precursors to the film is assumed to be the controlling factor. Others have assumed (55,56) that a chemical reaction is the controlling mechanism, and a polymerization mechanism similar to the Ziegler-Natta mechanism has been proposed. Both models work well within the data range. Based on commercial experience with a hydrocracked vacuum gas oil (57), it appears that more than polymerization or condensation occurs in a TLE, at least for this feedstock. By way of example typical outlet gas temperatures as a function of on-stream time are shown in Figure 4.

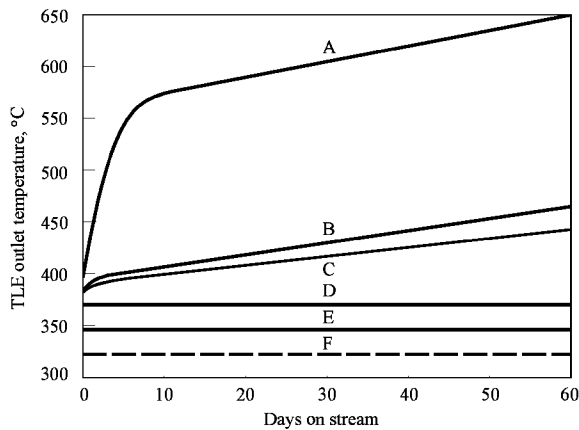


Fig. 4. Typical TLE outlet temperatures as a function of time on stream for various feedstocks: A, HVGO high severity; B, naphtha high severity; C, naphtha moderate severity; D, *n*-butane; E, ethane; and F, steam.

TLE fouling for gaseous feedstocks is different from that for liquid feedstocks. With gas cracking (especially with ethane), coke deposits on the TLE inlet tubesheet, but does not significantly build up inside the tubes. With time, enough coke builds up and tubes are partially or fully blocked and the heat-transfer surface is reduced. This gives rise to higher outlet temperatures and high pressure drop. Now, larger tubes are used to prevent plugging. Generally, only a small temperature rise is observed from start to end of run.

Only a few commercial vendors manufacture high pressure shell and tube transferline exchangers. Notably, Schmidtsche Heissdamp GmbH of Germany, Borsig of Germany, Mitsubishi of Japan, Mitsui of Japan, Babcock-Hitachi of Japan, and Struthers-Wells of the United States manufacture these TLEs. There are, however, many vendors who build low pressure TLEs.

Both Mitsubishi and Mitsui TLEs differ drastically from other designs. Mitsubishi offers a TLE with an integral steam drum and cyclone for vapor-liquid separation. The pyrolysis gas flows in the shell side, and is claimed to accomplish the decoking of the furnace and the transferline exchanger in one operation. The Mitsui quench cooler uses three concentric tubes as the tube element, and requires steam-air decoking to clean the TLE (58,59).

During radiant coil decoking, the TLE is also partially decoked (at least for ethane cracking, in which most of the coke is deposited on the inlet tubesheet). For complete decoking, the furnace is usually cooled down and the TLEs are separated from the coil and hydrojetted with high pressure water. In some cases, the coke in the TLEs can also be burnt off, and hence no mechanical cleaning is required (60). The radiant coil is always cleaned by burning the coke with steam and air in different proportions and at different temperatures for 12 to 48 hours. Since hydrogen burns off very rapidly, the initial concentration of oxygen is kept low to avoid temperature overshoot. Usually, all coils in a given heater are decoked, and the effluent of decoking is sent to a

decoking pot or to a firebox for burning. One patent claims single-coil decoking, while other coils in the same heater are cracking hydrocarbons which is true on-line decoking. Unfortunately, the decoking effluent contains CO, CO₂, and carbon particles that affect the downstream units, so widespread use of this method is limited. Usually the shorter the cracking time, the shorter is the decoking time for the same feedstock and cracking severity.

The inside of the convection tubes rarely foul, but occasionally the liquid unsaturates in feedstocks tend to polymerize and stick to the walls and thus reduce the heat transfer. This soft coke is normally removed by mechanical means. In limited cases, the coke can also be burnt off with air and steam. Normally, the outside surface of the convection section fouls due to dust and particles in the flue gas. Periodically (6 to 36 months), the outside surface is cleaned by steam lancing. With liquid fuel firing, the surface may require more frequent cleaning.

Recovery and Purification. The pyrolysis gas leaves the transferline exchanger at 300 to 400°C for gaseous and light naphtha feeds, and at 550 to 650°C for heavy liquid feeds. The lowest temperature is at the beginning of an operating cycle when the exchangers are clean. In order to minimize any further cracking for liquid feeds, the temperature must be quickly reduced. This is achieved by spraying quench oil directly into the effluent. For naphtha-based plants, quenching is performed before primary fractionation. In gas oil plants, quenching is done immediately after the transferline exchanger, resulting in two-phase flow in the transferline.

For all feeds the effluent is separated into desired products by compression in conjunction with condensation and fractionation at gradually lower temperatures.

Figure 5 represents a typical flow diagram of an ethylene plant for a naphtha feedstock. The quenched effluent enters the gasoline fractionator where the heavy fuel oil cuts are separated from the bulk of the effluent stream. The function of the fractionator is to make a sharp separation between the pyrolysis gasoline and pyrolysis fuel oil. The quench oil is cooled to 185°C by generating dilution or low pressure steam. The bottom temperature of the fractionator has to be critically controlled, since at high temperatures the quench material is unstable. Carbon (asphaltenes) is deposited, and the viscosity of the fluid increases. When light feeds are cracked, the fuel oil content is low, and therefore higher concentrations of unstable material are present. When gas oil is cracked, though the fuel oil content is high, it is mainly unconverted feed; hence the concentration of unstable material is low, and a higher bottom temperature can be tolerated.

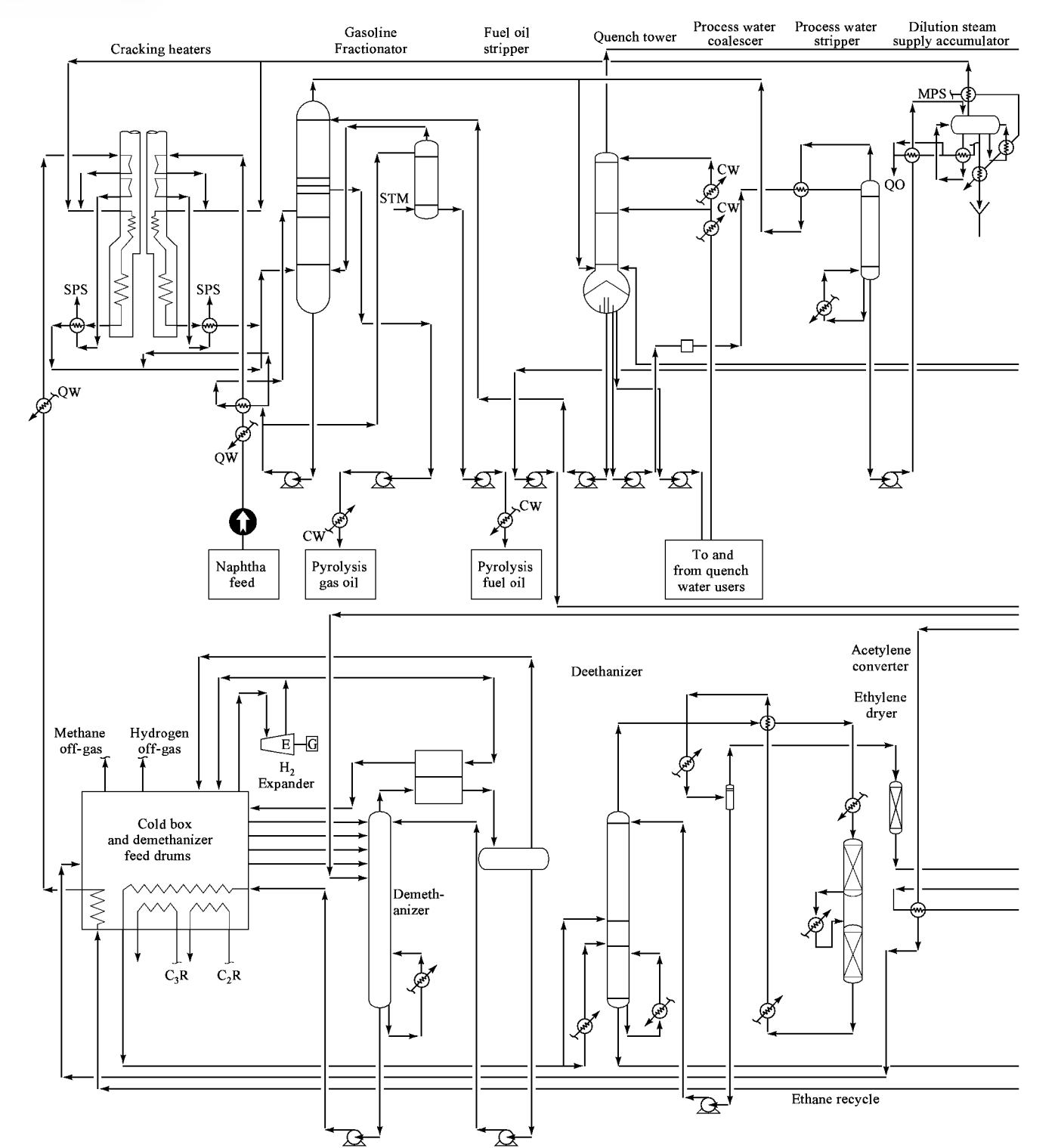


Fig. 5left. Schematic flow diagram of an ethylene plant using naphtha feedstock. CW = cooling water; QW = quench water; QO = quench oil; LPS = low pressure steam; MPS = medium pressure steam; SPS = super high pressure steam; C₃R = propylene refrigerant; and C₂R = ethylene refrigerant.

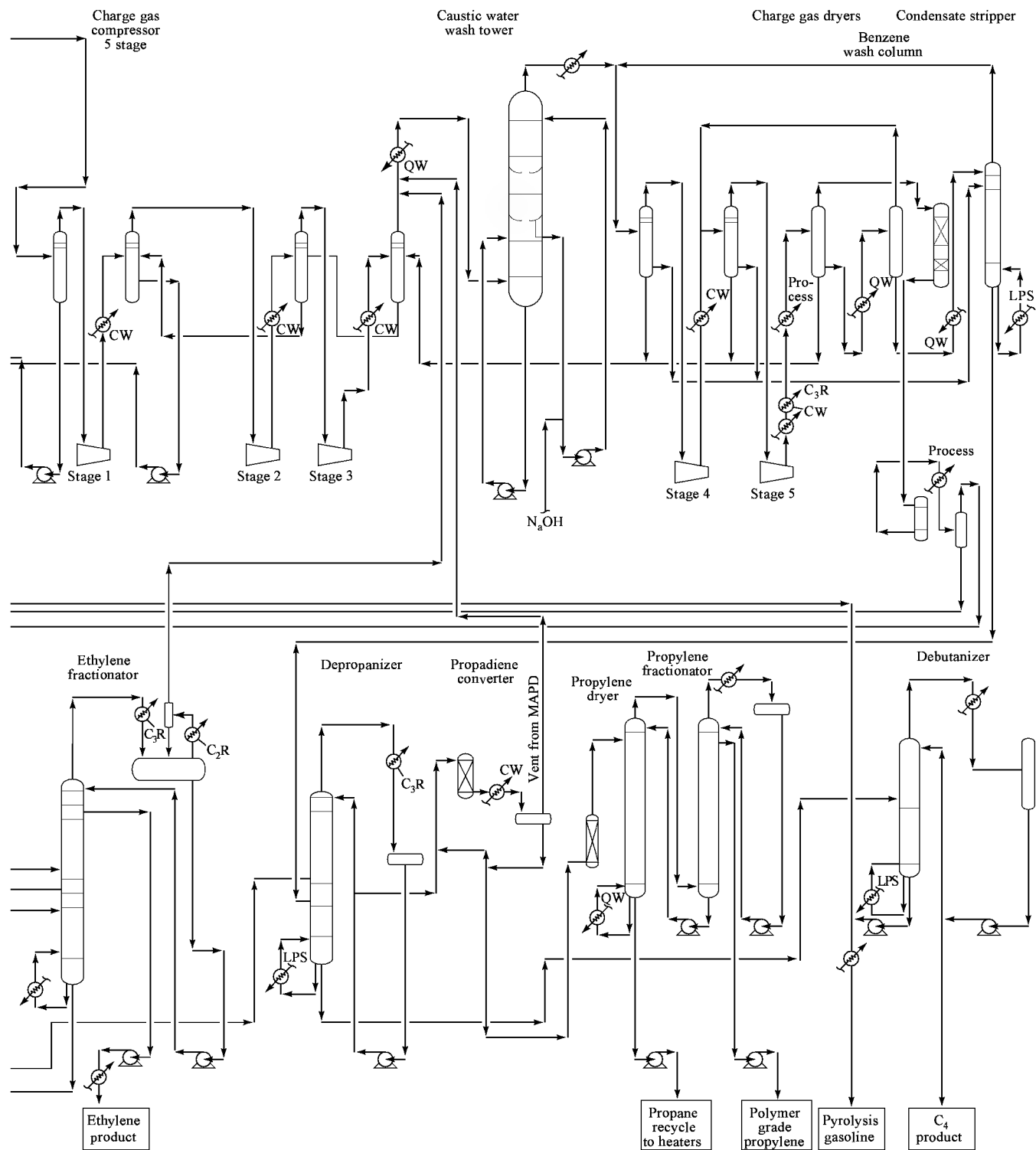


Fig. 5right. Continued

Because of the small temperature difference between the hot quench oil and the steam generators, and the large amount of heat to be removed from the pyrolysis gas, quench oil flow has to be large and is typically in the range of 15 to 25 times the flow of feedstock to the heaters. Because of coke particles in the quench oil, specially designed pumps are used. The dilution steam generator is one of the largest exchangers in the plant consisting of multiple shells. In addition, spare pumps and exchangers are needed because of severe fouling. This optimization problem has been analyzed (61,62) in detail.

The quench tower in essence operates as a partial condenser for the gasoline fractionator, condensing practically all of the steam and most of the pyrolysis gasoline components. In some designs, the quench tower and gasoline fractionator are combined into a single structure. Separation of the water phase from the gasoline phase occurs in a quench water drum. Hot quench water is used as a process heat source for the recovery section. A portion of the gasoline phase is refluxed for the gasoline fractionator, while the remainder is sent to the gasoline stripper for stabilization. This gasoline has an octane number (RON) of 95 to 99, and is usually blended with other gasoline products.

The pyrolysis gas leaving the quench tower is compressed to 3.5 MPa (35 atm) in a four- or five-stage centrifugal compressor. The number of stages is determined by the maximum temperature permissible for the material of construction and the fouling tendency of the pyrolysis gas. The compressor consists of two or three compressor casings driven by a single extraction condensing turbine. For large plants two turbines may be used. Water and hydrocarbons are separated from the pyrolysis gas between stages and recycled.

Acid gases (CO_2 and H_2S) are removed after the third or fourth compression stage. This is an optimum location since the gas volume has been reduced significantly in the previous stages and the acid gases have not contaminated any final products. When the sulfur content of the feed is low as for naphtha feeds, scrubbing with a dilute caustic soda solution is economical. Typically 4 to 12% free caustic solution is used. Relatively weak solutions are preferred to avoid the precipitation of sodium salts and to minimize the formation of sodium complexes and "yellow oil." The pyrolysis gas leaving the scrubber contains less than 1 ppm of acid gases, and is further treated by a water wash to remove any hydroxide carryover. A detailed analysis has been given (63). Plants designed to process high sulfur content feeds, eg, higher than 500 ppm, often contain a regenerative acid gas removal system upstream of the caustic scrubber. These systems may employ monoethanolamine, diethanolamine, or alkazid as solvents with a standard absorber-desorber design.

Disposal of the spent caustic solution can be a troublesome environmental problem. Depending on the plant location, acid gases are either sent to a fired heater or treated in a Claus unit for conversion of hydrogen sulfide to elemental sulfur.

Complete removal of water from the pyrolysis gas is achieved with molecular sieve dryers. Typically, there are two dryers; one is in normal operation while the other is being regenerated. The dryers are designed for 24 to 48 hours between successive regenerations and high pressure methane heated with steam at 225°C is the preferred regeneration medium. Activated alumina was used in older plants, but it is less selective than molecular sieves (qv).

After acid gas removal the pyrolysis gas from the last stage of compression is cooled by propylene refrigerant and sent to a condensate stripper. This tower separates the C₃ and heavier products, which exit the bottom, from the C₂ and lighter components.

The pyrolysis gas is partially condensed at essentially constant pressure over the stages of the cascade refrigeration system to about 165 °C where only the hydrogen remains in the vapor state. The stage condensates (only one is shown in Fig. 5) are fed to the appropriate trays of the demethanizer. Hydrogen (95 mol %) is withdrawn from the lowest temperature stage separator. The demethanizer is designed for complete separation of methane from ethylene and heavier components and operates at 0.7 MPa (7 atm). Condensing propylene refrigerant supplies heat to the reboilers and vaporizing refrigerant condenses the reflux. The demethanizer overhead consists of methane (95 mol %) with some minor impurities of hydrogen, carbon monoxide, and traces of ethylene. Brazed aluminum plate-fin exchangers are used for the multipass cryogenic heat-transfer services, and are installed in a cylindrical or rectangular carbon steel container, commonly known as a cold box. This unit is filled with perlite or rockwool for insulation.

The system just described is for the so-called low pressure demethanizer. Some licensors offer a high pressure demethanizer design (3.5 MPa), combined with either front-end deethanization or depropanization.

The demethanizer bottoms, consisting of ethylene and heavier components, are sent to the deethanizer, a conventional tray-type fractionator operating at a pressure of 2.4 to 2.8 MPa (~25 atm). An overhead stream containing C₂ hydrocarbons and a bottoms product of C₃ and heavier hydrocarbons are produced. Since acetylene is not usually recovered, the deethanizer overhead is heated to 20–100°C and hydrogen is added. The mixture is passed over a fixed bed of palladium catalyst. Because of the exothermicity of the acetylene hydrogenation, multiple beds with intermediate cooling are preferred. The acetylene hydrogenation reactor effluent contains less than 1 ppm of acetylene, but contains traces of methane and hydrogen. This is known as rear-end acetylene hydrogenation, and is preferred to front-end hydrogenation because of higher selectivity and precise temperature control.

Front-end hydrogenation is also possible. This approach uses a different type of catalyst, and the reactor is located upstream of the demethanizer. For this design, a deethanizer or depropanizer tower is located upstream of the demethanizer to remove heavy fractions. This approach has been utilized by some licensors with some success.

During acetylene hydrogenation, there is a net gain of ethylene since, under normal conditions, more acetylene is hydrogenated to ethylene than ethylene is hydrogenated to ethane. In most designs, traces of carbon monoxide are used to control the selectivity of hydrogenation. The catalyst is deactivated over time due to coke and green oil (a polymer formed as a result of side reactions) production, and is therefore regenerated periodically (3 to 6 months). The kinetics of this reaction has been discussed (64,65). New catalysts do not require CO addition.

When acetylene is recovered, absorption–desorption towers are used. In the first tower, acetylene is absorbed in acetone, dimethylformamide, or methylpyrrolidinone (66,67). In the second tower, absorbed ethylene and ethane are rejected. In the third tower, acetylene is desorbed. Since acetylene decomposition can result at certain conditions of temperature, pressure, and composition, for safety reasons, the design of this unit is critical. The handling of pure acetylene streams requires specific design considerations such as the use of flame arrestors.

After acetylene hydrogenation, the dried gas enters an ethylene–ethane separator known as an ethylene fractionator. This column contains 80 to 150 trays, and a reflux ratio of 2.5 to 4.0 is typical depending on feed composition. A pasteurizing section is usually provided at the top of the fractionator for removal of residual hydrogen, carbon monoxide, and methane. The conditions of the reboiler and overhead condenser are ideal for the use of a closed heat pump (68), with propylene refrigerant as the working fluid. Condensing refrigerant vapor supplies heat to the reboiler, and the refrigerant boiling under low pressure generates the cooling required in the overhead condenser. An open heat pump can be used with the integration of the ethylene–ethane fractionator into the ethylene refrigeration system. The ethane withdrawn from the tower bottom is recycled to the heaters and cracked to extinction.

The condensate stripper bottoms and the deethanizer bottoms are processed in the depropanizer for a sharp separation of C₃ hydrocarbons and C₄ and heavier hydrocarbons. The reboiler and tower internals are routinely fouled by rubberlike polymers, and require periodic mechanical cleaning. To minimize this problem, two types of designs are used. In the first design, the bottoms temperature is set low, which results in an operating pressure requiring propylene refrigerant instead of cooling water for the condensation of the overhead product. In the second design, cooling water is used, and a high bottom temperature is required. The latter requires a full spare tower and reboiler for continuous operation while the former requires only a spare reboiler. A two-tower system is also used to provide the benefits of both approaches, ie, low fouling and low refrigeration, but this option is somewhat more costly.

The depropanizer bottoms are further processed in the debutanizer for separation of C₄ product from light pyrolysis gasoline. The debutanizer operates at a moderate pressure of 0.4 to 0.5 MPa, and is a conventional fractionator with steam heated reboilers and water cooled condensers.

The overhead of the depropanizer is sent to the propylene fractionator. The methylacetylene (MA) and propadiene (PD) are usually hydrogenated before entering the tower. An MAPD converter is similar to an acetylene converter, but operates at a lower temperature and in the liquid phase. Due to recent advances in catalysis, the hydrogenation is performed at low temperatures (50–90°C) in trickle bed reactors (69). Only rarely are methylacetylene and propadiene recovered.

Because of the low relative volatility, fractionation of propylene and propane is even more difficult than the fractionation of ethylene and ethane. The propylene fractionator operates at a pressure of 1.8 to 2.0 MPa with more than 160 trays required for a high purity propylene product. Often a two-tower design is employed when polymer grade (99.5%+) is required. A pasteurization section may also be used when high purity is required. The bottoms product contains mainly propane that can be recycled to the cracking heaters or used as fuel. Typical tower dimensions and internals for a 450,000 t/yr ethylene plant with naphtha feed are summarized in Table 7.

Table 7. Characteristics of Towers in an Ethylene Plant^a

Service	Diameter, m	Height, m	Type, tray pack	Design temp, °C	Design pressure, kPa ^b
gasoline fractionator	7.1	35.6	pall ring	340	450
pyrolysis fuel oil stripper	1.5–3.2	14.3	valve	340	450
quench tower	7.2–14.4	32.8	packed	340	450
process water stripper	3.0–11.0	32.8	valve	210	450
caustic water wash tower	3.6	38	bubb pack	85	1270
condensate stripper	1.5	26.4	valve	95	1300
demethanizer ^c	2.0–3.0	64.7	pall ring	142	880
deethanizer	2.7	39.7	valve	45–85	2450
ethylene fractionator	3.7	84.3	valve	–45	2040
depropanizer	2.3	22–32	valve	100	2020
propylene fractionator	4.1	47.9	valve	75	2190
debutanizer	2.4	34.3	valve	140	760
acetylene absorber	2.1	51.9	valve	–35–110	2200
acetylene stabilizer	6.5–22.5	9–29	pall ring	35–110	1600

acetylene stripper	1.2–2.1	26.6	pall ring	200	1400
depentanizer	1.5	37.1	valve	150	450
BTX tower	2.6	23.2	valve	180	450
H ₂ S stripper	1.1	22	valve	195	9300

^a Ethylene capacity 450,000 t/yr; tower material is carbon steel.
^b To convert kPa to atm, divide by 101.3.
^c Tower material is stainless steel.

Energy Efficiency Improvement. Modern plants are more energy efficient than those designed in the 1980s. Reduction in energy consumption was made possible by improvements in cracking coil technology and improvements in recovery section design. Some improvements may not only reduce energy, but can also increase the capacity of an existing plant.

Quench Oil Viscosity Control. Increasing the bottoms temperature of the gasoline fractionator increases the viscosity, and flux oil is often added to reduce the viscosity. Effluents from ethane cracking can be used as a viscosity control stripping medium, allowing a temperature of 195°C in the gasoline fractionator bottom. This reduces the quench oil pumping power, eliminates the flux oil addition (69), and improves energy efficiency.

Feed Saturation. When gas feeds like ethane and propane are cracked, dilution steam can be added via direct humidification in towers known as feed saturators. This design reduces the load on the dilution steam system and or medium pressure (MP) steam level.

Predemethanization. The conventional design employs a single-step multiple feed tower. Utilization of a second tower upstream of the existing primary tower reduces the load on the primary tower. This reduces the propylene refrigeration power and reduces the propylene chilling loads. This approach is typically uneconomical with low pressure demethanizers, but has been combined with high pressure demethanizer systems by some licensors.

Demethanizer Overhead Expander and Multifeed Fractionation. Incorporation of an expander into the conventional high pressure demethanizer system eliminates bottlenecks in the refrigeration system, the demethanizer condenser, and charge gas compressor. It reduces the cost by lowering the refrigeration power. Multiple feed deethanization and ethylene fractionation debottlenecks the deethanizer, ethylene fractionator, and the refrigeration systems, thereby reducing power consumption.

Tower Internals and Equipment Modification. Tower capacity expansion can be achieved through the use of random or structured packing, or through the use of higher capacity trays such as the UOP multiple downcomer tray. Packing has been used in the gasoline fractionator, water quench tower, caustic and amine towers, demethanizer, the upper zone of the deethanizer, debutanizer, and condensate strippers. Packing reduces the pressure drop and increases the capacity.

Improved and redesigned rotors of modern compressors save considerable power. The ethylene fractionator and the propylene refrigeration condensers can be replaced with extended surface tube bundles instead of conventional tube bundles.

Dephlegmeters. Dephlegmeters accomplish feed gas separation while maintaining simplicity of process design, fabrication, and operation. Cryogenic dephlegmeters are brazed aluminum (plate-fin or core) heat exchangers specially designed to operate as mass-transfer devices. Depending on the feed composition, 5 to 15 theoretical stages can be obtained for one dephlegmator. Application of dephlegmeters to ethylene plants has been discussed (70,71).

Other Routes to Ethylene Production

In addition to conventional thermal cracking in tubular furnaces, other thermal methods and catalytic methods to produce ethylene have been developed. None of these are as yet commercialized.

Advanced Cracking Reactor. The selectivity to olefins is increased by reducing the residence time. This requires high temperature or reduction of the hydrocarbon partial pressure. An advanced cracking reactor (ACR) was developed jointly by Union Carbide with Kureha Chemical Industry and Chiyoda Chemical Construction Co. (72). A schematic of this reactor is shown in Figure 6. The key to this process is high temperature, short residence time, and low hydrocarbon partial pressure. Superheated steam is used as the heat carrier to provide the heat of reaction. The burning of fuel (H₂ and CH₄) with pure oxygen generates temperatures of 2000°C, and the cracking reaction is carried out at 950 to 1050°C (73–75) with a residence time of less than 10 milliseconds. Since the residence time in the reactor is so low, a specially designed Ozaki quench cooler for rapid quenching is required. A prototype was in operation for over 18 months during the 1980s. Unfortunately, all very high temperature processes produce high amounts of acetylene (>2 wt%). Acetylene hydrogenation is a significant cost factor if acetylene has no market.

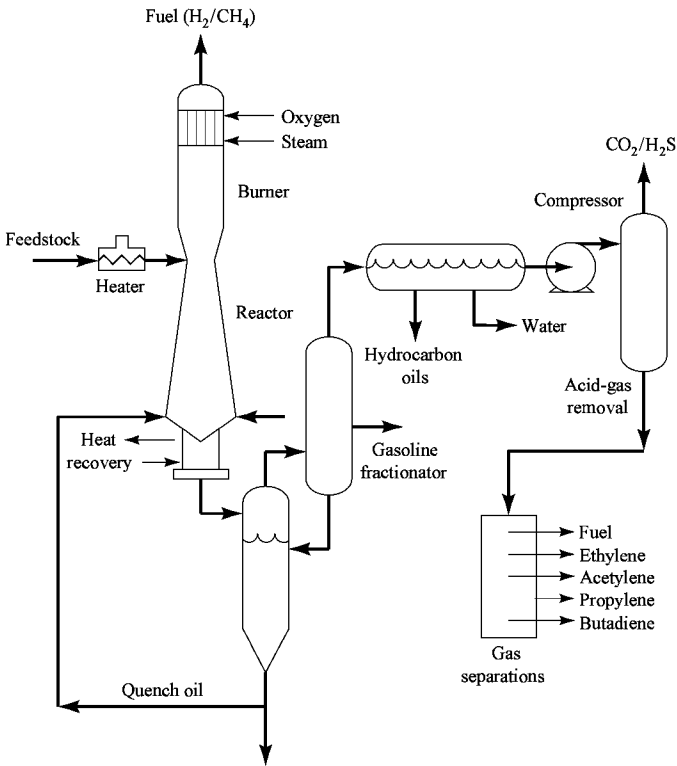


Fig. 6. Sketch of an advanced cracking reactor (73).

Adiabatic Cracking Reactor. This principle is based on the injection of hydrocarbon feedstock into the flue gas at elevated temperatures.

Because of the high initial temperature (1200°C), the feed can be instantaneously vaporized, and a very high rate of decomposition can be obtained. The temperature of the flue gas can be controlled by varying the oxygen–fuel ratio at the combustion chamber, and by the injection of steam in the combustion chamber. The endothermic nature of the cracking process causes the temperature to drop rapidly after the injection of the feed. A substantial increase (over 10 wt %) in olefin yield can be expected, but quenching the reaction to desired conditions is still a problem. In addition, the economics of the process are still not profitable. This route to ethylene production has been analyzed (76,77) using mathematical models. Instead of hydrocarbon, hydrogen has also been used as fuel that generates *in situ* dilution steam.

Fluidized-Bed Cracking. Since many fractions of crude oil are used individually as feedstocks in conventional cracking furnaces, logically many researchers aimed at cracking crude oil directly. This cannot be done in conventional coils because of severe fouling in the convection section and radiant coils, and in the transferline exchangers. Therefore, various types of fluid-bed processes have been developed. Notably, Lurgi developed the sand cracker (78) using sand as the heat carrier, whereas BASF used coke particles as the fluidizing medium (79). Ube (80) used inorganic oxide as a heat carrier, and the Kunugi and Kunii process (81) used fluidized bed with coke as a heat carrier. Thermal regenerative cracking jointly developed by Gulf Chemical (now Chevron) and Stone and Webster (82) uses solid heat carriers in the fluid bed. A semicommercial unit operated for over two years in Texas in the early 1980s. Other thermal process have also been discussed (83).

Catalytic Pyrolysis. This should not be confused with fluid catalytic cracking, which is used in petroleum refining (see CATALYSTS, REGENERATION). Catalytic pyrolysis is aimed at producing primarily ethylene. There are many patents and research articles covering the last 20 years (84–89). Catalytic research until 1988 has been summarized (86). Almost all catalysts produce higher amounts of CO and CO₂ than normally obtained with conventional pyrolysis. This indicates that the water gas reaction is also very active with these catalysts, and usually this leads to some deterioration of the olefin yield. Significant amounts of coke have been found in these catalysts, and thus there is a further reduction in olefin yield with on-stream time. Most of these catalysts are based on low surface area alumina catalysts (86). A notable exception is the catalyst developed in the former USSR (89). This catalyst primarily contains vanadium as the active material on pumice (89), and is claimed to produce low levels of carbon oxides.

Cracking temperatures are somewhat less than those observed with thermal pyrolysis. Most of these catalysts affect the initiation of pyrolysis reactions and increase the overall reaction rate of feed decomposition (85). Applicability of this process to ethane cracking is questionable since equilibrium of ethane to ethylene and hydrogen is not altered by a catalyst, and hence selectivity to olefins at lower catalyst temperatures may be inferior to that of conventional thermal cracking. Suitability of this process for heavy feeds like condensates and gas oils has yet to be demonstrated.

Membrane Reactor. Another area of current activity uses membranes in ethane dehydrogenation to shift the ethane to ethylene equilibrium. The use of membranes is not new, and has been used in many separation processes. However, these membranes, which are mostly biomembranes, are not suitable for dehydrogenation reactions that require high temperatures. Technology has improved to produce ceramic and other inorganic (90) membranes that can be used at high temperatures (600°C and above). In addition, the suitable catalysts can be coated without blocking the pores of the membrane. Therefore, catalyst-coated membranes can be used for reaction and separation.

As an example the use of ceramic membranes for ethane dehydrogenation has been discussed (91). The construction of a commercial reactor, however, is difficult, and a sweep gas is required to shift the product composition away from equilibrium values. The achievable conversion also depends on the permeability of the membrane. Figure 7 shows the equilibrium conversion and the conversion that can be obtained from a membrane reactor by selectively removing 80% of the hydrogen produced. Another way to use membranes is only for separation and not for reaction. In this method, a conventional, multiple, fixed-bed catalytic reactor is used for the dehydrogenation. After each bed, the hydrogen is partially separated using membranes to shift the equilibrium. Since separation is independent of reaction, reaction temperature can be optimized for superior performance. Both concepts have been proven in bench-scale units, but are yet to be demonstrated in commercial reactors.

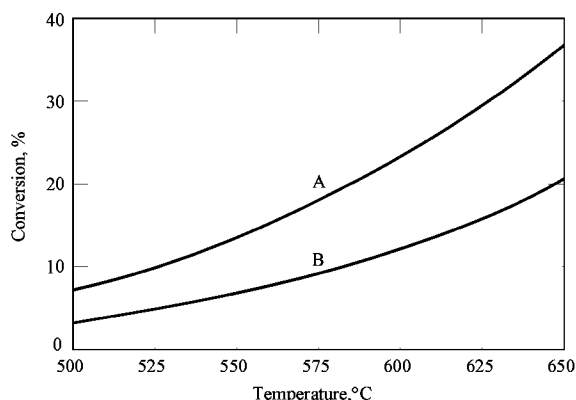
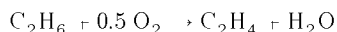


Fig. 7. Equilibrium conversion of ethane versus temperature at 210 kPa in a membrane reactor. The effect of hydrogen removal on ethane conversion is shown: A, 80% H₂ removal; B, 0% removal.

Dehydrogenation. The dehydrogenation of paraffins is equilibrium-limited and hence requires high temperatures. Using this approach and conventional separation methods, both Houdry and UOP have commercialized the dehydrogenation of propane to propylene (92). A similar concept is possible for ethane dehydrogenation, but an economically attractive commercial reactor has not been built.

Oxydehydrogenation. Because of the limitations of ethane dehydrogenation equilibrium, research has focused on ways to remove one of the products, namely hydrogen, by chemical methods. In this way, hydrogen is oxidized to water and there is no equilibrium limitation.



However, the same oxygen also oxidizes ethane and ethylene to CO₂ and other oxygenated products. Therefore, selectivity to olefins is a serious consideration. Literature citations (93,94) claim development of low temperature, highly selective oxydehydrogenation catalysts. Although this process has not yet been commercialized, it seems promising.

Oxidative Coupling of Methane. The most stable paraffin is methane, where it is difficult to break the C–H bond. During the 1980s, new catalysts were developed to activate methane, producing methyl radicals. These methyl radicals combine to give ethane, which further undergoes pyrolysis reactions. This has been discussed in detail (95,96). Recently, Eindhoven University Research Group (97) has proposed incentives for commercialization of this process. It has been shown that very high selectivity to ethylene can be obtained with lanthanum-based catalysts (98). The economics of this process have been discussed (95). According to these sources, this process is not economical when conventional feedstocks (naphtha, LPG, etc) are inexpensive. The process could be economical when methane is available in abundance at extremely low cost, such as in Saudi Arabia and other geographic locations. Since this process does not depend on crude oil for raw feed, research has continued in many countries, and it is possible that it may soon be commercialized.

Methanol to Ethylene. Methanol to ethylene economics track the economics of methane to ethylene. Methanol to gasoline has been fully developed and, during this development, specific catalysts to produce ethylene were discovered. The economics of this process have been discussed, and a catalyst (Ni SAPO 34) with almost 95% selectivity to ethylene has been claimed (99). Methanol is converted to dimethyl ether, which decomposes to ethylene and water; the method of preparation of the catalyst rather than the active ingredient of the catalyst has made the significant improvement in yield (100). By optimizing the catalyst and process conditions, it is claimed that yields of ethylene, propylene, or both are maximized. This is still in the bench-scale stage.

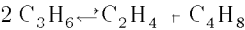
Ethanol to Ethylene. The economics of this process depend on the availability and price of ethanol. High volume production of ethylene

from ethanol, which is derived from fermentable raw materials, cannot normally compete with ethylene produced in large hydrocarbon-based olefin units. This process, however, offers several advantages to a country with abundant fermentation materials but limited hydrocarbon resources (102).

Activated alumina and phosphoric acid on a suitable support have become the choices for an industrial process. Zinc oxide with alumina has also been claimed to be a good catalyst. The actual mechanism of dehydration is not known. In industrial production, the ethylene yield is 94 to 99% of the theoretical value depending on the processing scheme. Traces of aldehyde, acids, higher hydrocarbons, and carbon oxides, as well as water, have to be removed. Fixed-bed processes developed at the beginning of this century have been commercialized in many countries, and small-scale industries are still in operation in Brazil and India. New fluid-bed processes have been developed to reduce the plant investment and operating costs (102,103). Commercially available processes include the Lummus processes (fixed and fluidized-bed processes), Halcon Scientific Design process, NIKK JGC process, and the Petrobras process. In all these processes, typical ethylene yield is between 94 and 99%.

Ethylene from Coal. There are several possible routes to ethylene from coal based on both conventional and emerging technologies. Synthesis gas derived from coal gasification can be converted to hydrocarbons by the Fischer-Tropsch (FT) process. The conventional FT process produces high molecular weight saturated and olefinic hydrocarbons and oxygenated compounds. Ethylene can be recovered directly or after pyrolyzing the ethane and naphtha produced. In general this is not an economical process; however, there is a plant based on this route in South Africa (104). Some of the new FT catalysts recently developed maximize ethylene and propylene yield (105). However, in Mobil's MTG (methanol to gasoline) or MTO (methanol to olefins), zeolite catalysts are used and high yields of olefins are reported (106).

Propylene Disproportionation. A commercial plant utilizing the disproportionation of propylene to ethylene was built in 1966 by Gulf Oil of Canada Ltd. utilizing technology developed by Phillips.



Since the above reaction is reversible, it can be used to produce either propylene or ethylene and butenes depending on relative prices. For example, when propylene is inexpensive, ethylene and butenes are produced. When ethylene is inexpensive and there is a high demand for propylene, ethylene is dimerized to butene and then the reverse reaction is utilized. A commercial plant based on the reverse reaction has been built on the Gulf Coast.

Ethylene as a By-Product. The contribution to world ethylene production is small, but not zero. In petroleum refining fluid catalytic cracking (FCC) units, small amounts of ethylene are produced but generally not recovered, except in a few locations where large FCC units are adjacent to petrochemical facilities.

Advanced Computer Control Systems and Training Simulators

An ethylene plant contains more than 300 equipment items. Traditionally, operators were trained at the site alongside experienced co-workers. With the advent of modern computers, the plant operation can be simulated on a real-time basis, and the results displayed on monitors (107). Computers are used in a modern plant to control the entire operation, eg, they are used to control the heaters and the recovery section (108). A well-controlled plant is much more profitable than a poorly controlled plant. For the heaters, a model-based control system is gaining importance (109). Instead of simply controlling the coil outlet temperature (COT), severity is actually controlled. The measurement of severity (either $\text{C}_3\text{H} / \text{C}_2\text{H}_4$ or $\text{C}_3\text{H} / \text{CH}_4$ ratio) requires on-line effluent analysis using chromatographs. However, these chromatographs have significant lag time. To overcome this lag time, sophisticated kinetic models are used to predict the severity for a given COT and compare it against the actual measurement as it becomes available. Although COT is used as the set point, severity is the control variable (108). This also provides a means to optimally adjust severity as fouling in the radiant coil and in the transfedine exchanger occurs.

Shipment and Storage

In 1989, over 500,000 t yr of ethylene were traded internationally. The principal exporting countries were in the Middle East, and the principal importing countries were in Western Europe and Asia Pacific. The tanker fleet that transported the ethylene numbered approximately 30 vessels with capacities ranging from 2000–6500 t (110). These tankers are of the semi-refrigerated type, and transport liquid ethylene at atmospheric pressure and –104°C. The tankers include reliquefaction plants on board since it is too expensive to vent ethylene. To accommodate the increase in international trade of ethylene, ethylene terminals have been built in the United States, Far East, Western Europe, and the Middle East with capacities of 35,000 t, 300,000 t, and 70,000 t, respectively (110).

The quantity of ethylene transported by international tankers accounts for only 1% of production. The majority of ethylene produced in the United States and Western Europe is moved by integrated pipeline systems.

In the United States, the Gulf Coast produces and consumes the majority of the U.S. ethylene production. The plants are located along the Texas southeast coast extending into Louisiana (111). The plants are served by a system of pipelines connecting the production and consuming plants. In 1989 there were 10 producers, 38 consumers, and 15 combined producers and consumers at the same site. The ethylene is transported as a gas at very high pressures of 5–5.7 MPa (49–56 atm) (112). Since the critical temperature for ethylene is 9.2°C, the ethylene remains as a gas at the elevated pressures. The pipelines are buried approximately 3–5 m below the ground so the surrounding ground temperature rarely approaches the critical temperature.

Western Europe also operates an integrated pipeline system. The system links the large producers and consumers in Germany, Holland, France, and Belgium. The pipeline system is being expanded both in terms of ethylene and the addition of propylene. Expansion of the pipeline is being instituted to phase out the use of rail tankers to transport the ethylene. Producers feel that pipeline transport offers superior specification guarantees that rail or barge tankers cannot offer since the operation is not directly under the user's control (113).

Very little data exists on the costs of transport via pipeline. In mid-1990, the cost of transport in Western Europe was approximately \$24 /t regardless of distance plus \$0.12 /km (113).

Inside operating plants a certain amount of storage is provided to smooth out operational upsets. The capacity stored can range from a few hours to a few days of production depending on the owner's operating philosophy. Small amounts of ethylene are stored as a liquid in pressure tanks and held at the required temperature by refrigeration available from the operating unit. This method of storage is only appropriate for a few hours of production, since the tanks must be expensive, heavy, thick-walled vessels. For larger amounts of ethylene, cryogenic low pressure storage is used. Cryogenic storage allows use of thinner-walled tanks that are less expensive (112). The ethylene is kept below its critical temperature, and therefore lower pressure can be used. To keep the ethylene below its boiling point (–103.7°C), some ethylene is vaporized and then passed to a refrigeration plant for recovery (112). Cryogenic storage tanks employ an outer and an inner wall with the interspace filled with an insulating material (114).

For storage of very large quantities, underground caverns or jugs leached out of underground salt domes are used. As the ethylene is pumped into the cavern it displaces a pool of brine. The pool of brine exerts pressure on the ethylene, keeping it in the cavern. The ethylene is removed by returning brine to the cavern. Since the caverns are located a few hundred meters below ground level, temperatures are fairly constant and always above the critical temperature of ethylene (112).

Specification and Analysis

Polyethylene is the predominant derivative of ethylene requiring very high purity (99.9% plus). A typical polymer-grade specification is given in Table 8. Almost all ethylene plants use gas chromatography or gas chromatography combined with mass spectrometry for analysis. For specific impurities like sulfur, water, and other hydrocarbons and elements, ASTM standard tests are recommended (D2504-67, D2505-67, 2785-70, E203-75).

Table 8. Specification for Polymer-Grade Ethylene

--	--

Component	Value
ethylene, mol % , min	99.95
methane + ethane	balance
other impurities, mol ppm, max	
hydrogen	5.0
acetylene	1.0
oxygen	1.0
carbon monoxide	1.0
carbon dioxide	1.0
propylene	10.0
C ₄ +	10.0
water	2.0
total sulfur	2.0
methanol	5.0
total chlorine	2.0
DMF	1.0
other compounds	5.0

Safety and Environmental Factors

Although ethylene is a colorless gas with a mild odor that is not irritating to the eyes or respiratory system, it is a hydrocarbon and therefore a flammable gas. All vessels must be designed for handling the liquids and gases during operation at the temperatures and pressures that exist, and safety and depressuring valves must be provided to relieve excessive pressure. Releasing of hydrocarbons into the air in large amounts must be avoided because of health and fire hazards. If hydrocarbons must be released into the air, it is done under a blanket of steam. To protect the plant and personnel in case of fire, a complete fire fighting system is provided with tanks grouped to minimize fire and provided with foam makers and deluge systems. Reviews at various stages of a project assure safety is given constant attention in the plant design (see HAZARD ANALYSIS AND RISK ASSESSMENT). An ethylene plant produces liquid, gaseous, and solid wastes that must be disposed of in an environmentally safe manner. Liquid wastes generated within the complex consist of wastewater streams of relatively low organic content, and process wastes of high organic content. Wastewater from various units and operations are segregated according to the wastewater characteristics, such as type of contaminants, concentration, special treatment, or pretreatment requirements. A segregated sewer system allows for the most efficient treatment of the wastewaters.

Atmospheric emissions from the facility are either controlled or fugitive in nature. Controlled emissions are released from process venting, waste incineration, decoking operations, and heater firing. Fugitive emissions occur from product loading and storage and equipment and valve leaks. In general, all continuous process vents are flared or combusted in the furnaces. If required, the process vents are scrubbed prior to flaring to minimize acid gas emissions. Flow monitors are installed in major branches in the flare collection header to monitor process venting. A smokeless flare has a normal destruction efficiency of over 98%.

Solid wastes are treated in a solid waste disposal area to reduce their volume and or toxicity prior to final disposal in a secured landfill. Combustible wastes can be incinerated in a slagging rotary kiln to reduce volume and toxicity.

Uses

Almost all ethylene produced is consumed as feedstock for manufacturing other petrochemicals. Only a very small amount has been used in the agricultural industry for ripening fruits. Table 9 lists the principal ethylene derivatives and capacities.

Table 9. Ethylene Derivatives and Capacities^a

Product	Actual		Projected avg growth, 1991–1994, %
	1989, 10 ³ t	1990, 10 ³ t	
polyethylene			
LDPE	16,361	16,440	0.9
LLDPE	6,213	7,123	12.6
HDPE	10,916	11,182	5.5
ethylene dichloride	9,602	10,094	3.3
ethylbenzene–styrene	4,295	4,539	6.5
ethylene oxide–glycol	7,892	8,162	6.3
ethanol	958	958	0.0
linear olefins–alcohol	2,064	2,236	4.4
vinyl acetate	1,156	1,178	0.7
other	2,599	2,658	0.4

^a Worldwide derivative capacity as ethylene equivalent in 1000 t/yr (115).

Although some ethylene is shipped across the oceans in large quantities, the preference is to ship first-generation products such as polyethylene, ethylbenzene, etc.

Economic Aspects

In 1989, world capacity for the production of ethylene was approximately 58 × 10⁶ t. The United States production capacity accounted for almost 30% of the world capacity, or approximately 17 × 10⁶ t, followed by Western Europe with almost 26% or 15 × 10⁶ t (116). Although the United States and Western Europe account for half of the world's production capacity, the most rapid growth in capacity has been occurring in the developing areas of the world. Table 10 shows the projected worldwide increase in capacity through 1993. Growth in capacity in the developing countries is almost twice that of North America and Western Europe. However, demand for ethylene is not expected to increase at the same high rate, leading to a shift in exports and imports. Table 10 also shows trade in ethylene in 1989 and projected trade in 1993.

Table 10. Projected Ethylene Capacity and Trade 1989–1993^a

	Net export or (import),
--	-------------------------

Region	1989 Capacity, 10 ³ t	1993 Capacity, 10 ³ t	Annual growth rate, %	10 ³ t	
				1989	1993
North America	20.6	26.1	4.9	53	80
Western Europe	15.1	19.0	4.7	(191)	(55)
Asia Pacific	10.0	15.6	9.3	(291)	(287)
Middle East	3.8	6.1	9.9	504	244
South America	2.2	3.4	9.1	(32)	(120)
Eastern Europe	6.7	7.2	1.5	(50)	(102)
<i>Total</i>	<i>58.4</i>	<i>77.4</i>	<i>5.8</i>		

^a Ref. 116.

Ethylene is first in total market value among petrochemicals. Based on 1989 production capacity of 58×10^6 t, the total potential market value would be approximately \$29 $\times 10^9$ based on an ethylene price of \$500/t.

Table 11 shows the historical sales price for ethylene and propylene in the United States. The economics for an ethylene plant are strongly affected by the prices for the propylene co-product. During most of the early 1980s, the price of ethylene was depressed, leading to a reluctance on the part of producers to install new capacity. By the late 1980s, the demand for ethylene had increased to such an extent that the existing capacity could only barely keep up with demand. Between 1987 and 1988 the price for ethylene almost doubled, encouraging producers to increase capacity. As the new capacity became available, the market price for ethylene began to drop. This trend of lower prices for ethylene will continue until surplus capacity is used up, forcing an increase in price.

Table 11. Prices^a for Ethylene and Propylene in the United States, \$/t

Year	Ethylene	Propylene
1981	570	500
1982	415	500
1983	440	415
1984	415	420
1985	330	370
1986	330	260
1987	330	360
1988	630	460
1989	660	450
1990	550	390
1991	500	420
1992	420	330

^a Refs. 111, 117, and 118.

The economics for the production of ethylene depends to a large extent on the prices for feedstocks and co-products. In the United States the feedstocks of choice have been the lighter feeds of ethane and propane, as opposed to Western Europe and the Far East which favor naphtha (116). Table 12 shows the percentage of ethylene produced from the various feedstocks in the United States since 1984. Approximately 70% of U.S. ethylene production is from ethane, propane, and butane. This percentage has remained fairly constant throughout the 1980s and it is assumed it will remain at this level. During the same time frame the use of naphtha as a feedstock has more than doubled, supplanting the use of gas oil as a feedstock.

Table 12. U.S. Ethylene Produced from Various Feedstocks, %^a

Year	Ethane propane butane	Naphtha	Gas oil
1984	70	10	20
1985	77	10	13
1986	71	16	13
1987	74	15	11
1988	70	25	5
1989	69	26	5
1990	68	27	5
1991	71	24	5
1992	71	24	5

^a Refs. 116, 118.

Ethane feed gives the lowest cost of production and the lowest capital investment. As the feeds become successively heavier, cost of production increases as well as the capital investment required. Depending on the cost of feedstock and the value of the co-products, processing heavier feedstocks can lead to lower returns on investment. Table 13 shows the effect on capital investment for various feedstocks as well as for a range of capacities.

Table 13. Ethylene Plant Investment and Cost of Production as a Function of Feedstock and Plant Capacity^a

Plant capacity, t/yr	Ethane	Propane	Butane	Naphtha	Gas oil
	<i>Relative plant investment</i>				
300,000	1.17	1.35	1.40	1.70	1.93
400,000	1.06	1.22	1.27	1.54	1.75
500,000	1.00	1.15	1.20	1.45	1.55
600,000	0.96	1.10	1.15	1.39	1.58
700,000	0.89	1.02	1.07	1.29	1.47
800,000	0.84	0.97	1.01	1.22	1.39
	<i>Relative cost of production</i>				
300,000	1.05	1.15	1.16	1.15	1.48
400,000	1.02	1.11	1.13	1.11	1.43
500,000	1.00	1.09	1.11	1.08	1.40

600,000	0.99	1.08	1.09	1.06	1.38
700,000	0.97	1.07	1.08	1.05	1.37
800,000	0.96	1.06	1.07	1.04	1.35

^a Factors are per ton of ethylene produced.

Another significant factor affecting capital investment is plant capacity. Prior to 1970, ethylene plant capacity was rarely greater than 300,000 t yr. During the 1970s and 1980s, technological advances resulted in plants being built with capacities of 500,000–700,000 t yr. The incentive to build these larger plants was the economic advantages of scale resulting from the reduction in capital requirements per ton of ethylene.

By 1990 it was theoretically possible to build a single-train (no duplication of compressors or other equipment except for cracking heaters) ethylene plant with a capacity of approximately 900,000 t yr. The limitation above this capacity becomes the suction volume to the charge gas compressor. There are reports of one plant operating at 950,000 t yr with as yet no limitation. Other factors limiting the size of ethylene plants are that larger equipment items require more costly field fabrication as opposed to shop fabrication, limiting potential savings in investment and increased risk resulting from large initial capital outlay and changing market conditions.

BIBLIOGRAPHY

"Ethylene" in *ECT* 1st ed., Vol. 5, pp. 880–898, by J. Happel, New York University, and E. I. Becker, Polytechnic Institute of Brooklyn; in *ECT* 2nd ed., Vol. 8, pp. 499–523, by D. L. Caldwell and I. Lichtenstein, The Lummus Co.; in *ECT* 3rd ed., Vol. 9, pp. 393–431, by L. Kniel, O. Winter, and C-H. Tsai, Combustion Engineering Inc.

1. K. M. de Reuck and co-workers, eds., *Ethylene (Ethene)-International Thermodynamic Tables of the Fluid State-10*, Blackwell Scientific Publishers, Oxford, UK, 1988.
2. *TRC Thermodynamic Tables*, Thermodynamic Research Center, The Texas A & M University System, College Station, Tex., 1986.
3. K. E. Starling, *Fluid Thermodynamic Properties for Light Hydrocarbon Systems*, Gulf Publishing Co., Houston, 1973.
4. D. R. Douslin and R. H. Harrison, *J. Chem. Thermodyn.* **8**, 301 (1976).
5. R. H. Harrison and D. R. Douslin, *J. Chem. Eng. Data* **22**, 24 (1971).
6. F. S. Bonscher, L. M. Shipman, and L. C. Yen, *Hydrocarbon Process.* **53** (1), 145 (1974).
7. N. B. Vargaftik, *Tables on the Thermophysical Properties of Liquids and Gases*, John Wiley & Sons, Inc., New York, 1975.
8. F. A. Carey, *Organic Chemistry*, McGraw-Hill Book Co., Inc., New York, 1987.
9. P. Sykes, *A Guide to Mechanism in Organic Chemistry*, 4th ed., Longman, London, 1975.
10. S. A. Miller, *Ethylene and Its Industrial Derivatives*, Baum, London, 1969.
11. H. Ulrich, *Raw Materials for Industrial Polymers*, Hanser Publishers, Munich, 1988.
12. K. H. Reichert and W. Geiseler, eds., *Polymer Reaction Engineering*, Hanser Publishers, Munich, 1983.
13. H. Sardina and co-workers, "New Styrene Plant Optimizer," paper presented at the *AIChE Meeting*, New Orleans, La., Mar. 1988.
14. T. Horigome, R. Johnson, and D. Gandhi, "The First World Scale Plant using the Lummus Unocal UOP Process for Ethylbenzene is a Huge Success," paper presented at the *Catalyst Conference*, Muenchen, July 4, 1991.
15. P. J. Lewis and F. G. Dwyer, *Oil & Gas J.* 55 (Sept. 26, 1977).
16. B. C. Gates, J. R. Katzer, and C. G. A. Schuit, *Chemistry of Catalytic Processes*, McGraw-Hill Book Co., Inc., New York, 1979.
17. L. Kniel, O. Winter, and K. Stork, *Ethylene-Keystone to the Petrochemical Industry*, Marcel Dekker, Inc., New York, 1980.
18. G. F. Froment, *Chem. Eng. Sci.* **36**, 1271 (1981).
19. A. L. Albright, W. L. Crynes, and W. H. Corcoran, *Pyrolysis Theory and Industrial Practice*, Academic Press, Inc., New York, 1983.
20. F. O. Rice, *J. Am. Chem. Soc.* **53**, 1959 (1931); **55**, 3035 (1933).
21. K. M. Sundaram and G. F. Froment, *Ind. Eng. Chem. Fund.* **17**, 174 (1978).
22. D. L. Allara and D. A. Edelson, *Int. J. Chem. Kinet.* **7**, 479 (1975).
23. D. L. Allara and R. A. Shaw, *J. Phys. Chem. Ref. Data* **9**, 523 (1980).
24. M. Dente and E. Ranzi, in L. F. Albright, B. L. Crynes, and W. H. Corcoran, eds., *Pyrolysis Theory and Industrial Practice*, Academic Press, Inc., New York, 1983.
25. C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971.
26. M. Boudart, *Kinetics of Chemical Processes*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1968.
27. N. N. Semenov, *Some Problems in Chemical Kinetics and Reactivity*, trans. by M. Boudart, Princeton University Press, N.J., 1959.
28. K. M. Sundaram and G. F. Froment, *Int. J. Chem Kinet.* **10**, 1189 (1978).
29. K. M. Sundaram and J. M. Fernandez-Baujin, *AIChE J.* **34**, 321 (1988).
30. C. E. Van Camp and co-workers, *Ind. Eng. Chem. Process Des. Dev.* **24**, 561 (1985).
31. S. B. Zdonik and E. J. Green, "Manufacturing Ethylene," reprinted from the *Oil Gas J. (1966–1970)*, Petroleum Publishing Co., Tulsa, Okla., 1970.
32. B. Lohr and W. Schwab, *Linde Rep. Sci. Technol.* **30**, 18 (1979).
33. L. L. Ross and W. R. Shu, *Adv. Chem. Ser.* **183**, 129 (1979).
34. J. V. Albano, T. Fukushima, and E. F. Olszewski, "Gas Turbine Integration in Ethylene Plants," paper presented at *ACHEMA 91, International Meeting on Chemical Engineering and Biotechnology*, June 12, 1991.
35. D. H. Cooke and W. D. Parizot, *Trans. ASME* **113**, 212 (1991).
36. F. P. Boer and co-workers, *CHEMTECH*, 312 (May 1990).
37. K. J. Laidler, *Chemical Kinetics*, McGraw-Hill Book Co., Inc., New York, 1965.
38. A. G. Buekens and G. F. Froment, *Ind. Eng. Chem. Process Des. Dev.* **7**, 435 (1968).
39. J. M. Fernandez-Baujin and S. M. Solomon, "An Industrial Application of Pyrolysis Technology: Lummus SRT-III Module," paper presented at the *First Chemical Congress of the North American Continent*, Mexico City, Nov. 30–Dec. 5, 1975.
40. K. M. Sundaram and G. F. Froment, *Chem. Eng. Sci.* **32**, 601 (1977); **32**, 609 (1977).
41. M. Hirato and S. Yosida, *Int. Chem. Eng.* **13**, 347 (1973).
42. P. Kumar and D. Kunzru, *Ind. Eng. Chem. Process Des. Dev.* **24**, 774 (1985).
43. L. P. Hillewaert, J. L. Dierickx, and G. F. Froment, *AIChE J.* **34**, 17 (1988).
44. S. W. Benson, *Thermochemical Kinetics*, McGraw-Hill Book Co., Inc., New York, 1968.
45. A. F. Trotman-Dickenson, *Adv. Free Radical Chem.* **1**, 1 (1965).
46. R. C. West, ed., *CRC Handbook of Chemistry and Physics*, CRC Press, Inc., Boca Raton, Fla., 1987.
47. J. A. Dean, *Lange's Handbook of Chemistry*, 13th ed., McGraw-Hill Book Co., Inc., New York, 1985.
48. I. Lichtenstein, *Chem. Eng. Prog.* **60**(12), 64 (1964).
49. A. G. Goossens, M. Dente, and E. Ranzi, *Process Control and Optimization Handbook*, p. 30.
50. P. M. Plehiers, G. C. Reyniers, and G. F. Froment, *Ind. Eng. Chem. Res.* **29**, 637 (1990).
51. K. M. Sundaram, P. S. Van Damme, and G. F. Froment, *AIChE J.* **27**, 946(1981).
52. D. S. Newsome and H. P. Leftin, *Chem. Eng. Prog.* **76**, (May 1980).

53. D. L. Trimm and C. J. Turner, *J. Chem. Technol. Biotechnol.* **31**, 195 (1981).
54. J. Chen and W. Vogel, "Fouling of Transferline Exchangers in Ethylene Services," paper presented at *74th National Meeting, AIChE*, New Orleans, March 1973.
55. M. Dente and co-workers, "Ethylene Cracker Transferline Exchanger Fouling," paper presented at *AIChE Meeting*, Houston, March 1983.
56. M. Dente and co-workers, "TLE Fouling Model," paper presented at *AIChE Spring National Meeting*, Orlando, Fla., Mar. 18–22, 1990.
57. K. M. Sundaram and J. M. Fernandez-Baujin, "An Experience on HVGO," paper presented at *AIChE Spring National Meeting*, Orlando, Fla., Mar. 18–22, 1990.
58. T. Sato and T. Katada, *Chem. Econ. Eng. Rev.* **4**(7), 14 (1972).
59. K. Sato, *Chem. Econ. Eng. Rev.* **9**(6), 16 (1977).
60. **Eur. Pat. 81,101,665.8** and **Ger. Pat. Appl. 010,000** (Mar. 15, 1980), (to BASF).
61. M. Picciotti, *Hydrocarbon Process.* **56**(5), 183 (1977).
62. *Ibid.*, **56**(6), 163 (1977); **56**(9), 179 (1977).
63. M. Raab, *Linde Rep. Sci. Technol.* **23**, 23 (1976).
64. W. K. Lam, "Acetylene Hydrogenation," paper presented at *Pacific Area Chemical Engineering Congress*, Acapulco, Mexico, Oct. 19–22, 1988.
65. L. Cider and N. H. Schoon, *Ind. Eng. Chem. Res.* **30**, 1437 (1991).
66. **U.S. Pat. 3,816,976** (June 18, 1974), K. Stork, L. Kniel, and S. Markbreiter (to Lummus).
67. U. Lorber, H. Reimann, and F. Rottmayr, *Chem. Eng.* **78**(7), 83 (1971).
68. A. Meili, *Chem. Eng. Prog.*, (June 6, 1990).
69. S. J. Stanley and R. M. Venner, "Revamp of World Scale Ethylene Unit," paper presented at *Fourth World Congress of Chemical Engineers*, Karlsruhe, Germany, June 16–20, 1991.
70. C. P. Bowen, "Olefin Expansion Strategies," paper presented at *Symposium on Advanced Petrochemical Technologies and High Tech Products*, Nijing, China, 1991.
71. J. A. Nachenberg, "ARS Update Revamp Concepts," paper presented at *Eighth Ethylene Forum*, The Woodlands, Tex., May 15–17, 1991.
72. T. Wett, *Oil Gas J.*, (Feb. 7, 1972).
73. R. L. Baldwin and G. R. Kamm, *Hydrocarbon Process.*, 127 (Nov. 1982).
74. J. D. Kearns, D. Milks, and G. R. Kamm, *Adv. Chem. Ser.* **183**, 107 (1979).
75. T. Hosoi and R. G. Keister, *Chem. Eng. Prog.* **78**, 11 (1975).
76. M. Dente, E. Ranzi, and S. Barendregt, "Adiabatic Cracking Yields Theoretically Predicted," paper presented at *AIChE Meeting*, New Orleans, July 1981.
77. M. Dente and co-workers, "Steam Cracking of Heavy Liquid Feedstocks," paper presented at *AIChE Spring National Meeting*, New Orleans, Nov. 1985.
78. P. Schmalfeld, *Hydrocarbon Process.*, (July 1963).
79. A. Steinhofner, O. Frey, and H. Nonnenmacher, *Hydrocarbon Process.*, (July 1963).
80. T. Matsunami and co-workers, *Hydrocarbon Process.*, 121 (Nov. 1970).
81. D. Kunugi, *Chem. Eng. Sci.* **35**, 1887 (1980).
82. A. F. Ellis, R. J. Gartside, and C. F. Bown, "TRC-A New Olefins Technology," paper presented at *90th AIChE Meeting*, Houston, 1981.
83. Y. C. Hu, *Hydrocarbon Process.*, 109, (Nov. 1982).
84. K. Kikuchi and co-workers, *Chem. Eng. Prog.*, 54 (June 1985).
85. A. A. Lemonidou and I. A. Vasalos, *Appl. Catal.* **54**, 119 (1989).
86. A. A. Lemonidou and co-workers, *Ind. Eng. Chem. Res.* **28**, 524 (1989).
87. A. A. Lemonidou and co-workers, "Catalyst Evaluation and Kinetic Study for Ethylene Production," paper presented at *AIChE Spring National Meeting*, Houston, Mar. 29–Apr. 2, 1987.
88. J. H. Kolts and G. A. Delzer, *Science* **232**, 744 (1986).
89. S. P. Chernykh, ed., *New Organic Synthesis Processes*, trans. by G. Leib, Mir Publishers, Moscow, 1991.
90. H. P. Hsieh, *Catal. Rev.-Sci. Eng.* **33**(1 & 2), 1 (1991).
91. A. M. Champagne, T. T. Tsotsis, and R. G. Minet, *Chem. Eng. Sci.* **45**, 2423 (1990).
92. B. V. Vora, P. R. Pujado, and R. F. Anderson, "Olefex—C₂-C₅ Dehydrogenation Updated," paper presented at *AIChE Meeting on New Developments in Olefin Technology*, New Orleans, Apr. 6–10, 1986.
93. S. J. Conway and J. H. Lunsford, *J. Catalysis* **131**, 513 (1991).
94. T. Laegreid, "ADH: The Dehydrogenation Process, Status, and Possibilities," paper presented at *SPUNG Seminar*, Norway, Sept. 24–25, 1991.
95. U. Preuss and M. Baerns, *Chem. Eng. Technol.* **10**, 297 (1987).
96. M. Baerns, "Progress in Methane Conversion Science—Technology—Economics," paper presented at *SPUNG Seminar*, Norway, Sept. 24–25, 1991.
97. M. Campman and co-workers, *Ethylene from Natural Gas—Proven and New Technologies*, M. Campman and co-workers, Eindhoven University of Technology, the Netherlands, 1990.
98. P. Pereira and co-workers, *Catal. Lett.* **6**, 255 (1990).
99. T. Inui and T. Takegami, *Hydrocarbon Process.*, 117 (Nov. 1982).
100. T. Inui, S. Phatanasri, and H. Matsuda, *Catalyst Sci. Technol.* **1**, 85 (1991).
101. M. E. Winfield, *Catalysts*, Vol. 7, Reinhold Publishing Co., New York, 1960, p. 93.
102. O. Winter, "Ethylene from Ethanol," paper presented at *First Brazilian Petrochemical Congress*, Rio de Janeiro, Nov. 7–12, 1976.
103. U. Tsao and J. W. Reilly, *Hydrocarbon Process.*, (Feb. 1978).
104. M. E. Dry, *Catalysis Sci. Tech.* **1**, 159 (1981).
105. B. Buessemeier, C. D. Frohning, and B. Cornils, *Hydrocarbon Process.* **55**(11), 105 (1976).
106. T. Inui, "Structure-Reactivity Relationship in Methanol to Olefin Conversion on Various Microporous Crystalline Catalysts," paper presented at *Division of Petroleum Chemistry, ACS Meeting*, Boston, Apr. 22–27, 1990.
107. T. Sinhora and S. Kasada, "A Real Time Dynamic Simulator for the Residue FCC Process," paper presented at *Eastern Simulation Conference*, Orlando, Fla., Apr. 18–21, 1988.
108. B. Stancato and co-workers, "Advanced Control Optimization for Mitsubishi Kashima Ethylene Plant," paper presented at *AIChE Spring Meeting*, Houston, April 7–11, 1991.
109. L. Kane, ed., *Advanced Process Control Handbook*, *Hydrocarbon Process.*, Sept. 1991.
110. Private communication, Marublni (Japan), 1988.
111. Chemical Market Associates Inc. (CMAI), *1989 World Light Olefins Analysis*, Houston, Tex., Aug. 1989.
112. D. L. Burdick and W. L. Leffer, *Petrochemicals for the Non-Technical Person*, Pennwell Books, Tulsa, Okla., 1983, pp. 53–56.
113. *Eur. Chem. News*, 13 (July 23, 1990).
114. R. H. Perry and C. H. Chilton, *Chemical Engineers Handbook*, 5th ed., McGraw-Hill Book Co., Inc., 1973, Chapt. 6, p. 88.
115. *1989–1990 Ethylene Annual*, Dewitt & Co. Inc., Houston, Tex.
116. *1989–1990 Ethylene Annual*, Dewitt & Co. Inc., Houston, Tex.
117. *CMAI Petrochemical Handbook*, 1988.

118. *CMAI Monomers Report*, 1989 through Aug. 1991.

K. M. Sundaram
M. M. Shreehan
E. F. Olszewski
ABB Lummus Crest, Inc.

Tempera-ture, °C	Vapor pres-sure, ^a kPa ^b	Enthalpy, J g ^{a,c}		Den-sity , ^d kg L	Heat capa-city, ^d J (kg·K) ^c	Surface tension, ^e mN m (dyn cm)	Viscosity, ^e mPa·s (cP)	Thermal conduc-tivity, W (m·K) ^f
		Liquid	Vaporiza-tion					
40	8.35	0	628.6	0.9488	1878	34.2	0.495	0.20
30	15.05							
20	25.73	38.8	605.4	0.9232	1912	30.9	0.400	0.18
10	42.00							
0	65.82	77.3	581.7	0.8969	1954	27.6	0.325	0.16
10	99.54							
20	145.8	115.3	557.3	0.8697	2008	24.3	0.26	0.15
30	207.7							
40	288.4	153.2	532.1	0.8413	2092	21.0 ^g	0.23	0.14
50	391.7							
60	521.2	191.8	505.7	0.8108	2247	17.8 ^g	0.20	0.14
70	681.0							
80	875.4	232.6	477.4	0.7794	2426	14.7 ^g	0.17	0.14
90	1108.7							
100	1385.4	277.8	445.5	0.7443	2782	11.8 ^g	0.16	0.13
120	2088	330.4	407.5	0.7052	3293 ^h	8.9 ^g		
140	3020	393.5	359.4	0.6609	4225 ^h	6.2 ^g		
160	4224	469.2	297.1	0.608		3.6 ^g		
180	5741	551.2	222.5	0.533				
195.8	7191							

^a Ref. 9.

^b To convert kPa to mm Hg, multiply by 7.50.

^c To convert J to cal, divide by 4.184.

^d Refs. 18,19.

^e Refs. 18–20.

^f Calculated, Ref. 23.

^g Calculated, Ref. 22.

^h Calculated, Ref. 21.

Table 3. Physical Properties of Ethylene Oxide Vapor from 298 to 800 K

Temperature, K	Entropy, ^{a,b} J (mol·K) ^c	Heat of formation, ^{b,d} kJ mol ^c	Free energy of formation, ^{a,b} kJ mol ^c	Viscosity, ^e μPa·s ^f	Thermal conductivity, ^g W (m·K)	Heat capacity, ^h J (mol·K) ^c
298	242.4	52.63	13.10			48.28
300	242.8	52.72	12.84	9.0	0.012	48.53
400	258.7	56.53	1.05	13.5	0.025	61.71
500	274.0	59.62	15.82	15.4	0.038 ⁱ	75.44
600	288.8	62.13	31.13	18.2	0.056 ⁱ	86.27
700	302.8	64.10	46.86	20.9	0.075 ⁱ	95.31
800	316.0	65.61	62.80		0.090 ⁱ	102.9

^a Ref. 21.

^b Calculated, Ref. 24.

^c To convert J to cal, divide by 4.184.

^d Refs. 12, 24.

^e Refs. 18–20.

^f μPa·s = 10^{−5} P.

^g Refs. 18,19.

^h Ref. 25.

ⁱ Calculated, Ref. 21.

Table 4. Physical Properties of Aqueous Solutions of Ethylene Oxide

Ethylene oxide		Specific gravity at 10 °C ^a	Freezing point ^b , °C	Boiling point ^a , °C
Wt %	Mol %			
0	0	1.000	0.0	100
2.5	1.0	0.9986	0.9	70
5	2.1	0.9977	1.6 (eutectic)	58
10	4.4	0.9944	5.6	42.5
15	6.7	0.9888	8.9	38
20	9.3	0.9816	10.4	32
30	14.9	0.9658	11.1 (max)	27

40	21.4	0.9500	10.4	21
50	29.0	0.9356	9.3	19
60	38.0	0.9227	7.8	16
70	48.8	0.9111	6.0	15
80	62.1	0.9005	3.7	13
90	78.6	0.8910	0.0	12
100	100	0.8828	111.7	10.4

^a Ref. 9.

^b Refs. 9 and 20.

Table 5. Flashpoints of Ethylene Oxide/Water Solutions ^a

Ethylene oxide concentration		Flash point, °C ^b	
Vol %	Wt %	Closed cup	Open cup
1.2	1.07	26	nd
2.0	1.80	16	nd
2.5	2.22	11	nd
3.0	2.66	5	nd
3.5	3.11	1	nd
4.0	3.53	nd	60
4.5	4.00	nd	43
5.0	4.43	nd	37
6.0	5.32	nd	18

^a Ref. 9.

^b nd not determined.

Table 6. Solubility of Gases in Ethylene Oxide, Henry's Constants, MPa ^{a,b}

Temperature, °C	Nitrogen	Argon	Methane	Ethane
0	289	171	63	8.6
25	221	144	62	11.0
50	189	131	62	13.3

^a To convert MPa to atm, divide by 0.101.

^b Ref. 26.

Table 7. Solubility of Ethylene Oxide in Water, mL Vapor ^a/mL Solvent ^b

Pressure, kPa ^c	Temperature, °C		
	5°C	10°C	20°C
20	45	33	20
27	60	46	29
40	105	76	49
53	162	120	74
67	240	178	101
80		294	134
93			170
101			195

^a Reduced to 0°C and 101.3 kPa ^c.

^b Ref. 27.

^c To convert kPa to mm Hg, multiply by 7.50.

Structure. The strained configuration of ethylene oxide has been a subject for bonding and molecular orbital studies. Valence bond and early molecular orbital studies have been reviewed (28). Intermediate neglect of differential overlap (INDO) and localized molecular orbital (LMO) calculations have also been performed (29–31). The LMO bond density maps show that the bond density is strongly polarized toward the oxygen atom (30). Maximum bond density lies outside of the CCO triangle, as suggested by the bent bonds of valence–bond theory (32). The ¹H-nmr spectrum of ethylene oxide is consistent with these calculations (33).

The structural parameters of ethylene oxide have been determined by microwave spectroscopy (34). Bond distances in nm determined are as follows: C–C, 0.1466; C–H, 0.1085; and C–O, 0.1431. The HCH bond angle is 116.6°, and the COC angle 61.64°. Recent *ab initio* studies using SCF, MP2, and CISD have predicted bond lengths that are very close to the experimental values (35,36).

Clathrate Formation. Ethylene oxide forms a stable clathrate with water (20). It is nonstoichiometric, with 6.38 to 6.80 molecules of ethylene oxide to 46 molecules of water in the unit cell (37). The maximum observed melting point is 11.1°C. An x-ray structure of the clathrate revealed that it is a type I gas hydrate, with six equivalent tetrakaidecahedral (14-sided) cavities fully occupied by ethylene oxide, and two dodecahedral cavities 20–34% occupied (38).

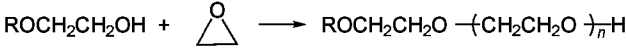
Chemical Properties

Ethylene oxide is a highly reactive compound, and so is used industrially as an intermediate for many chemical products. The three-membered ring is opened in most of its reactions. These reactions are very exothermic because of the tremendous ring strain in ethylene oxide, which has been calculated (39). Reviews of ethylene oxide reactions are given in References 40 and 41.

Polymerization. The reaction of ethylene oxide with a nucleophile introduces the hydroxyethyl group:



The product of this reaction can also react with ethylene oxide; if this process is repeated many times, a polymer is formed:



Low molecular weight polymers of ethylene oxide, poly(ethylene glycol), are formed by allowing ethylene oxide to react with water or alcohols under the proper conditions (see POLYETHERS). The average molecular weight can be varied from 200 to 14,000 (32,33).

Polymers with much higher average molecular weights, from 90,000 to 4×10^6 , are formed by a process of coordinate anionic polymerization (43–45). The patent literature describes numerous organometallic compounds, alkaline-earth compounds, and mixtures as polymerization catalysts. Iron oxides that accumulate in ethylene oxide storage vessels also catalyze polymerization. This leads to the formation of nonvolatile residue (NVR); no inhibitor has been found (46).

Crown Ethers. Ethylene oxide forms cyclic oligomers (crown ethers) in the presence of fluorinated Lewis acids such as boron trifluoride, phosphorus pentafluoride, or antimony pentafluoride. Hydrogen fluoride is the preferred catalyst (47). The presence of BF₃, PF₅, or SbF₅ salts of alkali, alkaline earth, or transition metals directs the oligomerization to the cyclic tetramer, 1,4,7,10-tetraoxacyclododecane [294-93-9] (12-crown-4), pentamer, 1,4,7,10,13-pentaoxacyclopentadecane [33100-27-6] (15-crown-6), and hexamer, 1,4,7,10,13,16-hexaoxacyclooctadecane [17455-13-9] (18-crown-6) by a template effect. Each cation maximizes the formation of the crown ether that best encircles its ionic radius (48,49).

Other Chemical Reactions.

With Water. Wurtz was the first to obtain ethylene glycol by heating ethylene oxide and water in a sealed tube (1). Later, it was noted that by-products, namely diethylene and triethylene glycol, were also formed in this reaction (50). This was the first synthesis of polymeric compounds of well-defined structure. Hydration is slow at ambient temperatures and neutral conditions, but is much faster with either acid or base catalysis (Table 8). The type of anion in the catalyzing acid is relatively unimportant (58) (see GLYCOLS).

Table 8. Rate Constants for the Hydrolysis of Ethylene Oxide^a

Temperature, °C	Acidic, <i>k</i> _a , L (mol·min) ^b	Neutral, ^c 10 ⁴ <i>k</i> _w , min ⁻¹	Basic, 10 ² <i>k</i> _b , L (mol·min) ^b
20	0.32	0.22	0.34
30	1.00	0.55	1.0
40	2.5	1.9	3.06
60		11.9	17.0
80		60.6	77

^a Refs.

Extrapolated to zero salt concentration.

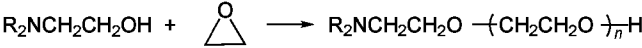
^c Rate increases dramatically above 100°C. *k*_w = 510, 1170, and 1730 × 10⁴ min⁻¹ at 113, 123, and 131°C, respectively.

With Alcohols. These reactions parallel those of ethylene oxide with water. The primary products are monoethers of ethylene glycol; secondary products are monoethers of poly(ethylene glycol) (42). Most are appreciably water-soluble.

With Organic Acids and Anhydrides. The carboxyl group of an organic acid reacts with ethylene oxide to give the corresponding ethylene glycol monoester. This product can also react with ethylene oxide to yield a poly(ethylene glycol) ester, or with another acid to produce a glycol diester. Ethylene glycol diesters may be obtained directly by the reaction of ethylene oxide with the acid anhydride.

With Ammonia and Amines. Ethylene oxide reacts with ammonia to form a mixture of mono-, di-, and triethanolamines. Nitrogen is a stronger nucleophile than oxygen (59). A small amount of water is essential for the reaction (60).

Complex nitrogen compounds are formed from the reaction of alkylamines with ethylene oxide (61). Thus diethylamine and ethylene oxide react to yield diethylaminoethanol. The dialkylaminoethanols can react with ethylene oxide to give amino poly(ethylene glycols):



Primary and secondary aromatic amines react with ethylene oxide to give the corresponding arylaminoethanols.

With Hydrogen Sulfide and Mercaptans. Ethylene oxide reacts with hydrogen sulfide to yield 2-mercaptoethanol and thiodiglycol (bis-2-hydroxyethyl sulfide) (62,63). Reaction conditions determine the proportions of each derivative. Three moles of ethylene oxide react with one mole of hydrogen sulfide in water to give the strong base tris(hydroxyethyl)sulfonium hydroxide, (HOCH₂CH₂)₃S⁺OH⁻; The reaction of ethylene oxide with long-chain alkyl mercaptans yields polyoxyethylene mercaptans, some of which are nonionic surfactants (64).

p-Oxathiane and *p*-dithiane are formed from ethylene oxide and hydrogen sulfide at 200°C in the presence of an aluminum oxide catalyst (65).

With Grignard Reagents. Ethylene oxide reacts with Grignard reagents (66), RMgX, to yield the corresponding two carbon homologue, RCH₂CH₂OH (67–69).

With Acyl Halides, Hydrogen Halides, and Metallic Halides. Ethylene oxide reacts with acetyl chloride at slightly elevated temperatures in the presence of hydrogen chloride to give the acetate of ethylene chlorohydrin (70). Hydrogen halides react to form the corresponding halohydrins (71). Aqueous solutions of ethylene oxide and a metallic halide can result in the precipitation of the metal hydroxide (72,73). The halides of aluminum, chromium, iron, thorium, and zinc in dilute solution react with ethylene oxide to form sols or gels of the metal oxide hydrates and ethylene halohydrin (74).

Phosphorus oxychloride reacts with ethylene oxide in the presence of aluminum chloride to give tris-2-chloroethyl phosphate, a valuable plasticizer (75). Phosgene reacts with ethylene oxide and other alkylene oxides to form esters of chlorocarbonic acid (76) (see CARBONIC AND CARBONOCHLORIDIC ESTERS).

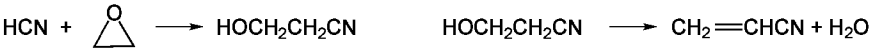
With Compounds Containing Active Methylene or Methine Groups. Compounds containing active –CH₂– or –CH– groups, such as malonic and monosubstituted malonic esters, ethyl cyanoacetate, and β-keto esters, react with ethylene oxide under basic conditions (72,77). Ethylene oxide and diethyl malonate react in the presence of sodium ethoxide to give diethyl (2-hydroxyethyl)malonate, which cyclizes to form α-carboethoxy-γ-butyrolactone.

The sodium salt of ethyl acetoacetate in ethanol at 0°C reacts with ethylene oxide to give 2-acetyl-4-butyrolactone, an intermediate for vitamin B₁ and anti-malarials (75).

With Phenols. The 2-hydroxyethyl aryl ethers are prepared from the reaction of ethylene oxide with phenols at elevated temperatures and pressures (78,79). 2-Phenoxyethyl alcohol is a perfume fixative. The water-soluble alkylphenol ethers of the higher poly(ethylene glycol)s are important surface-active agents. They are made by adding ethylene oxide to the alkylphenol at ca 200°C and 200–250 kPa (>2 atm), using sodium acetate or

hydroxide as a catalyst. The properties of these alkylphenol ethers can be varied over a wide range of solubility and performance characteristics by changing the alkyl chain(s) and the number of ethylene oxides added (80).

With Hydrogen Cyanide. Ethylene oxide reacts readily with hydrogen cyanide in the presence of alkaline catalysts, such as diethylamine, to give ethylene cyanohydrin. This product is easily dehydrated to give acrylonitrile in 80–90% yield:

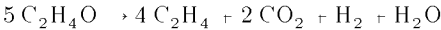


Ethylene cyanohydrin can be hydrolyzed to acrylic acid or esterified to give the corresponding alkyl acrylates (81) (see CYANOHYDRINS).

Miscellaneous Reactions. Ethylene oxide is considered an environmental pollutant. A study has determined the half-life of ethylene oxide in the atmosphere (82,83). Autodecomposition of ethylene oxide vapor occurs at ~500°C at 101.3 kPa (1 atm) to give methane, carbon monoxide, hydrogen, and ethane (84–86).

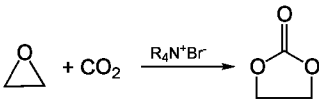
Isomerization of ethylene oxide to acetaldehyde occurs at elevated temperatures in the presence of catalysts such as activated alumina, phosphoric acid, and metallic phosphates (75). Iron oxides also catalyze this reaction. Acetaldehyde may be found as a trace impurity in ethylene oxide.

Disproportionation of ethylene oxide to ethylene and carbon dioxide also occurs at temperatures higher than about 150°C in the presence of high surface area iron oxides (87).



Diborane reacts with ethylene oxide at -80°C to form diethoxyborane and a solid polymer containing approximately eight ethylene oxide units per molecule (88). Potassium thiocyanate or thiourea react in aqueous solution with ethylene oxide to give ethylene sulfide (89).

Ethylene carbonate (1,3-dioxolan-2-one) is commercially prepared from ethylene oxide by the addition of carbon dioxide to ethylene oxide with either ammonium or alkali metal salts as catalysts (87):



Carbon disulfide reacts with ethylene oxide to give ethylene trithiocarbonate (90), and isocyanates yield derivatives of 2-oxazolidinone (91).

Ethylene oxide reacts with phosphonium halides to give ylides, which are used to synthesize olefins from carbonyl compounds, such as aldehydes and ketones (92).

Many other reactions of ethylene oxide are only of laboratory significance. These include nucleophilic additions of amides, alkali metal organic compounds, and pyridinyl alcohols (93), and electrophilic reactions with orthoformates, acetals, titanium tetrachloride, sulfenyl chlorides, halo-silanes, and dinitrogen tetroxide (94).

Manufacture

Ethylene oxide has been produced commercially by two basic routes: the ethylene chlorohydrin and direct oxidation processes. The chlorohydrin process was first introduced during World War I in Germany by Badische Anilin-und Soda-Fabrik (BASF) and others (95). The process involves the reaction of ethylene with hypochlorous acid followed by dehydrochlorination of the resulting chlorohydrin with lime to produce ethylene oxide and calcium chloride. Union Carbide Corp. was the first to commercialize this process in the United States in 1925. The chlorohydrin process is not economically competitive, and was quickly replaced by the direct oxidation process as the dominant technology. At the present time, all the ethylene oxide production in the world is achieved by the direct oxidation process.

The direct oxidation technology, as the name implies, utilizes the catalytic oxidation of ethylene with oxygen over a silver-based catalyst to yield ethylene oxide. The process can be divided into two categories depending on the source of the oxidizing agent: the air-based process and the oxygen-based process. In the first, air or air enriched with oxygen is fed directly to the system. In the second, a high purity oxygen stream (>95 mol %) from an air separation unit is employed as the source of the oxidizing agent. Union Carbide Corp. was the first to commercialize an air-based direct oxidation process in 1937. The first oxygen-based system was commercialized by Shell Oil Co. in 1958 (96).

Several companies have developed technologies for direct oxidation plants. Union Carbide Corp. and Dow Chemical use their own technologies. Shell Development, Scientific Design (96), and more recently, Nippon Shokubai, license ethylene oxide technology, and over 70% of present world capacity is based on their processes. Shell and Nippon Shokubai technologies are solely oxygen-based, and Scientific Design offers both air- and oxygen-based processes. All the ethylene oxide plants that have been built during the last 15 years were oxygen-based processes, and a number of existing ethylene oxide plants were converted from the air to the oxygen-based process during the same period (97). Extensive information on the early developments of the chlorohydrin and direct oxidation processes are reported in Reference 98. The total world production capacity of ethylene oxide in 1992 was about 9.6 × 10⁶ metric tons. Most ethylene oxide is consumed by its producers in making derivatives (99).

There are 12 producers of ethylene oxide in the United States. Table 9 shows the plant locations, estimated capacities, and types of processes employed. The total U.S. production capacity for 1992 was ca 3.4 × 10⁶ metric tons. The percentages of total domestic production made by the air- and oxygen-based processes are ca 20 and 80%, respectively. The largest producer is Union Carbide Corp. with approximately one-third of the United States' ethylene oxide capacity. About 94% of domestic ethylene oxide capacity is located on the Gulf Coast near secure and plentiful ethylene supplies. Plans for additional U.S. production in the 1990s have been announced by Union Carbide (incremental expansions), Formosa Plastics (at Pt. Comfort, Texas), and Shell (at Geismar, Louisiana) (101).

Table 9. Domestic Producers, Capacities, Process Types and Technology Used for Ethylene Oxide^a

Producer	Location	Capacity, 10 ³ t/yr	Process oxidant	Technology
BASF	Geismar, La.	218	oxygen	Shell
Dow	Plaquemine, La.	227	oxygen	Dow
Eastman	Longview, Tex.	91	oxygen	Shell
Hoechst-Celanese	Clear Lake, Tex.	209	oxygen	Shell
Olin	Brandenburg, Ky.	50	oxygen	Shell
Oxy Petrochemicals	Bayport, Tex.	250	oxygen	Shell
PD Glycols ^b	Beaumont, Tex.	202	oxygen	Scientific Design
Quantum	Morris, Ill.	113	oxygen	Scientific Design
Shell	Geismar, La.	364	oxygen	Shell
Sun Refining	Claymont, Del.	45	oxygen	Shell
Texaco	Port Neches, Tex.	332	oxygen	Scientific Design
Union Carbide	Seadrift, Tex.	349	air	Union Carbide
Union Carbide	Taft, La.	668	air, oxygen	Union Carbide

Total	3118
-------	------

^a Ref. 100 and miscellaneous sources.

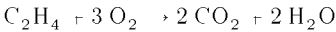
^b Joint venture: Oxychem and Du Pont.

Direct Oxidation Processes. The phenomenal growth in United States and world ethylene oxide production capacity since 1940 and the marked trend toward larger single-train plants is chiefly due to the great commercial success of the direct oxidation process. Compared to the chlorohydrin process, direct oxidation eliminates the need for large volumes of chlorine. Also, there are no chlorinated hydrocarbon by-products to be sold, processing facilities can be made simpler, and operating costs are lower (102). The main disadvantage of the direct oxidation process is the lower yield or selectivity of ethylene oxide per unit of feed ethylene consumed. The main inefficiency in the process results from the loss of ca 20–25% of the ethylene to carbon dioxide and water. Consequently, operating conditions must be carefully controlled to maximize selectivity.

All ethylene oxide direct-oxidation plants are based on the original process chemistry discovered by Lefort in 1931 (7,8). The main reaction is as follows:



The only significant by-products are carbon dioxide and water, which are formed either by complete combustion of ethylene:

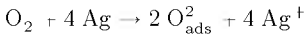


or by further oxidation of ethylene oxide:

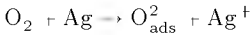


To prevent further oxidation of ethylene oxide, the ethylene conversion of the commercial processes is typically between 10 and 20%.

Although these reactions have been researched extensively and are the subjects of numerous patents, the precise reaction mechanism is not fully understood. The controversy has mostly centered on the nature of the oxygen species responsible for ethylene oxide formation (103). The results of various surface characterization studies indicate that there are at least three types of adsorbed oxygen species on silver: monoatomic chemisorbed oxygen, diatomic (molecular) oxygen, and subsurface oxygen. The first results from a dissociative adsorption of oxygen on a silver surface:

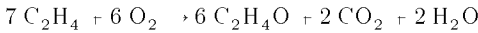


The second is nondissociative, and is more weakly bonded:



The third case arises when the temperature is higher than 420 K, at which point diffusion of atomic adsorbed oxygen from surface to subsurface region becomes appreciable (103).

During the 1970s and early 1980s, the prevailing theory was that ethylene reacts with diatomic oxygen to form ethylene oxide, leaving one oxygen atom on the surface (103). Monoatomic chemisorbed oxygen reacts with ethylene to form carbon dioxide and water. Since O_{ads}^2 must be removed from the silver surface before more O_{ads}^2 can form, the two reactions must be combined. The stoichiometry for the complete reaction is:



This leads to a limiting ethylene selectivity of 6/7 or 85.7%, which has been exceeded, as reported in several patents (104–106).

More recent studies provided the evidence for a different theory (107). The results of these studies indicate that monoatomic chemisorbed oxygen leads both to ethylene oxide formation, and to carbon dioxide and water formation. The controlling factor is the charge state of adsorbed monoatomic oxygen. Strongly negative-charged monoatomic oxygen acts as a base, leading to the abstraction of hydrogen from ethylene and complete combustion. When subsurface oxygen is present, it competes with adsorbed atomic oxygen for silver electrons, reducing the negative charge on the adsorbed oxygen. This increases its affinity toward the electron-rich double bond of ethylene. The molecular framework of ethylene is preserved, and the reaction path to ethylene oxide becomes predominant. This theory has gradually gained support and now guides the thinking in ethylene oxide catalyst research.

In addition to ethylene oxide, carbon dioxide, and water, small quantities of acetaldehyde and traces of formaldehyde are also produced in the process. They generally total less than 0.2% of the ethylene oxide formed. Acetaldehyde is most likely formed by isomerization of ethylene oxide, whereas formaldehyde is most likely formed by direct oxidation of ethylene (108).

A large amount of heat is released by the ethylene oxidation reactions. At 600 K, each kg of ethylene converted to ethylene oxide releases 3.756 MJ (3564 Btu); each kg of ethylene converted to carbon dioxide and water releases 50.68 MJ (48,083 Btu). Energy recovery and integration is a prime concern in process design (108).

Commercial processes operate under recycle conditions in a packed-bed, multitubular reactor. Reaction temperatures of 200–300°C are typical, and operating pressures of 1–3 MPa (10–30 atm) have been reported (96,102,109). The reactor is of the shell and tube type comprised of several thousand mild steel or stainless steel tubes, 20–50 mm inside diameter (96). The reactor can be either oil or boiling water cooled. Figure 1 is a schematic diagram of an oil-cooled reactor. Based on published information regarding catalyst productivities and space velocities, the reactor tube lengths are 6–12 m (96,102,109–114). These tubes are filled with a silver-based catalyst ca 3–10 mm dia supported on a carrier material with a surface area usually <1 m²/g (96,102,106,107). The yield (moles of product produced per moles of ethylene consumed in the process) is normally 70–80% depending on catalyst type, per pass conversion, reactor design, and a large number of process operating variables (97).

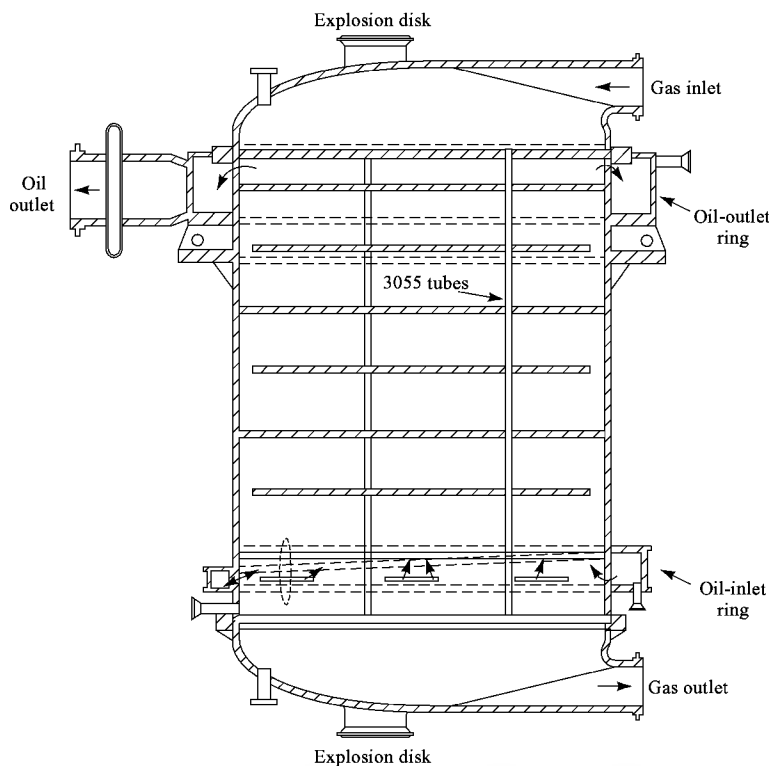


Fig. 1. Oil-cooled reactor for the oxidation of ethylene to ethylene oxide.

Technological innovations in catalyst development and process design and engineering have enabled ethylene oxide manufacturers to meet the commercial needs for larger facilities without using a great number of reactors (102,115). For example, in the Shell oxygen-based process, it is claimed that individual reactor capacities have increased from a modest 9,000 t yr in 1958 to 25,000 t yr in 1968, to ca 150,000 t yr in recent designs (115). There is a pronounced trend in both the United States and Europe toward larger single-train plant sizes. In the late 1950s, a 30,000 t yr ethylene oxide unit was considered large, whereas 10 years later plant sizes of 100,000–150,000 t yr were typical (116). Today some producers have plant capacities in excess of 250,000 t yr.

Air-Based Direct Oxidation Process. A schematic flow diagram of the air-based ethylene oxide process is shown in Figure 2. Published information on the detailed evolution of commercial ethylene oxide processes is very scanty, and Figure 2 does not necessarily correspond to the actual equipment or process employed in any modern ethylene oxide plant. Precise information regarding process technology is proprietary. However, Figure 2 does illustrate all the salient concepts involved in the manufacturing process. The process can be conveniently divided into three primary sections: reaction system, oxide recovery, and oxide purification.

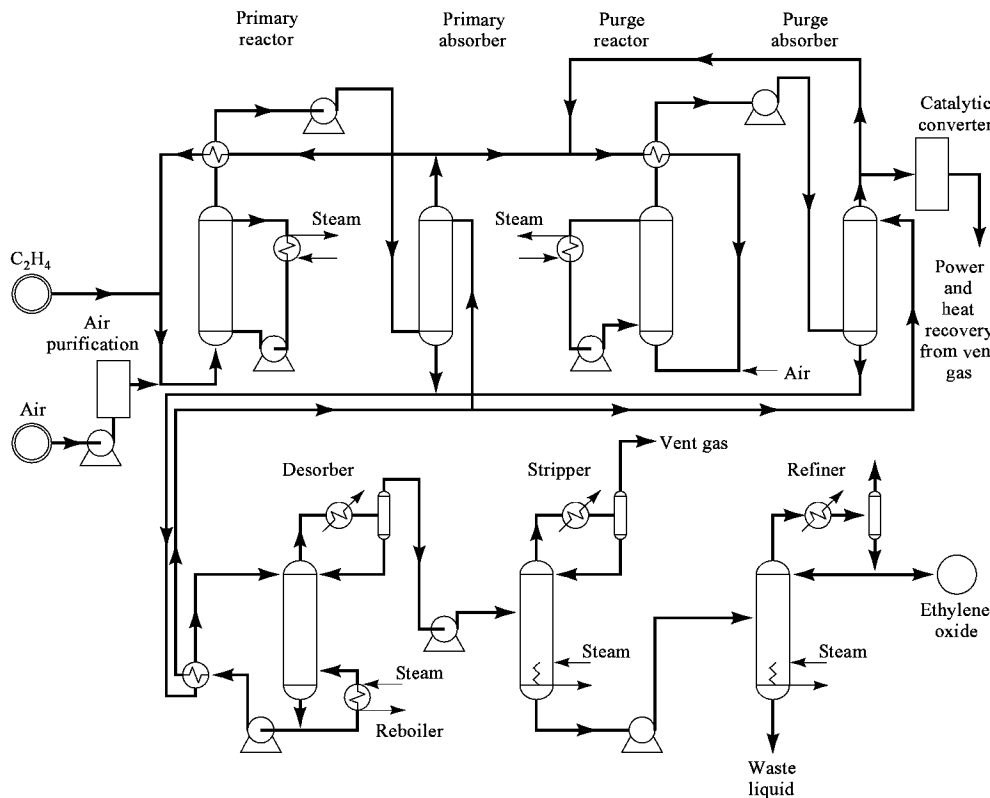


Fig. 2. Air-based direct oxidation process for ethylene oxide (96,102,109,117–119)).

In the first section, compressed air is filtered, purified (if necessary), and fed separately with ethylene into a recycle gas stream. This recycle stream feeds a bank of one or more primary multitubular reactors that operate in parallel. The number of primary reactors used depends chiefly on the plant capacity, size of the individual reactors, and the activity and the selectivity of the catalyst used. The ethylene is oxidized to ethylene oxide, carbon dioxide, and water in the packed-bed converters, and the heat of reaction is removed by circulating or boiling an organic oil on the shell side (108) (see HEAT EXCHANGE TECHNOLOGY), eg, Dowtherm, Tetralin, or other high boiling materials. The hot oil is cooled in a steam generator, producing considerable

amounts of high pressure steam for the ethylene oxide and other processes at the plant site (96,109).

The per pass ethylene conversion in the primary reactors is maintained at 20–30% in order to ensure catalyst selectivities of 70–80%. Vapor-phase oxidation inhibitors such as ethylene dichloride or vinyl chloride or other halogenated compounds are added to the inlet of the reactors in ppm concentrations to retard carbon dioxide formation (107,120,121). The process stream exiting the reactor may contain 1–3 mol % ethylene oxide. This hot effluent gas is then cooled in a shell-and-tube heat exchanger to around 35–40°C by using the cold recycle reactor feed stream gas from the primary absorber. The cooled crude product gas is then compressed in a centrifugal blower before entering the primary absorber.

The second important step of the process is ethylene oxide recovery from the crude product gas. This is accomplished in the primary absorber by countercurrent scrubbing with cold water in a column ca 18–20 m high. The ethylene oxide produced in the reactor is dissolved in the absorber water along with some nitrogen and carbon dioxide, and traces of ethylene, ethane, and aldehydes (102). The aqueous stream is removed from the base of the absorber and sent to a desorber. The unabsorbed gas from the main absorber overhead is split into two portions. The largest portion is recycled to the primary reactor after it cools the hot product gas in the shell-and-tube heat exchanger, and the circulation cycle is repeated. A much smaller fraction of the primary adsorber overhead gas stream is heat-exchanged to raise its temperature, and it is then fed as the main stream to the secondary or purge reactor system. In the purge reactor, more air may be added to increase the oxygen content of the feed gas. The gases leaving the purge reactor are heat-exchanged against the feed gas to the same reactor, and then enter a purge absorber. In the purge absorber, ethylene oxide is removed with water in the same manner as in the main absorber.

The chief purpose of the purge reactor system is to allow reaction of a substantial portion of the ethylene content of the purge gas, which must be vented from the main reactor system in order to prevent accumulation of inert gases, primarily nitrogen and carbon dioxide (109). Figure 2 shows a two-stage, air-based plant with a single-purge reactor. In larger plants, three or more stages of reaction may be used to improve overall yield of product (109,113). In such cases the flow scheme is virtually the same, except additional purge reactor–absorber systems are added in series to the first purge reactor. The scrubbed gas from the last purge absorber may be partly recycled to the same purge reactor inlet or vented from the system.

In some cases, the ethylene content of the vent gas leaving the last purge reactor makes it economical to further process this gas for energy recovery (109). Such a scheme not only extracts valuable power from the vent gas, but also reduces considerably the hydrocarbon emissions from the process. Several such schemes have been described in References 108 and 122–125. The basic scheme involves heating the ethylene lean gas (<1.5 mol %) to ca 200°C, and then passing it into a catalytic combustion chamber filled with an active oxidation catalyst containing a noble metal, eg, platinum. Such an active catalyst should burn virtually all of the ethylene and ethane in the vent gas, thereby raising its temperature to 400–600°C. The hot, pressurized gas expands in a turbine that is coupled to the feed gas compressor in the main process. The hot exhaust gases from the turbine are used to generate steam. Using such a scheme, a 10% reduction in the overall cost is claimed for the manufacture of ethylene oxide by the air-based process (125).

The third key section of the process deals with ethylene oxide purification. In this section of the process, a variety of column sequences have been practiced. The scheme shown in Figure 2 is typical. The ethylene oxide-rich water streams from both the main and purge absorbers are combined, and after heat exchange are fed to the top section of a desorber where the absorbate is steam stripped. The lean water from the lower section of the desorber is virtually free of oxide, and is recirculated to the main and purge absorbers. The concentrated ethylene oxide vapor overhead is fed to the ensuing stripper for further purification. If the desorber is operated under vacuum, a compressor is required.

The ethylene oxide recovered in the desorber contains some carbon dioxide, nitrogen, aldehydes, and traces of ethylene and ethane. In the stripper the light gases are separated overhead and vented, and the partially purified ethylene oxide is sent from the bottom of the stripper to the mid-section of a final refining column. The ethylene oxide from the refining section should have a purity of >99.5 mol %. The final product is usually stored as a liquid under an inert atmosphere.

The overall economics of the process are strongly dictated by the design of the reaction system and the actual operating conditions used. The catalyst properties, as they influence reactor design and operating variables are, therefore, of the greatest significance. Specific information on actual conditions employed in the manufacture is not disclosed. However, the general ranges suggested by literature and patent reviews are summarized in Table 10.

Table 10. Ranges of Reaction System Variables in the Direct Oxidation Process for Ethylene Oxide ^a

Variable	Air oxidation	Oxygen oxidation
ethylene, mol %	2–10	20–35
oxygen, mol %	4–8	4–8
carbon dioxide, mol %	5–10	5–10
ethane, mol %	0–1.0	0–1.0
temperature, °C	220–277	220–235
pressure, MPa ^b	1–3	2–3
space velocity ^c , h ^{–1}	2000–4500	2000–4500
pressure drop, kPa ^d	41–152	41–152
conversion, %	20–65	8–12
selectivity or yield (mol basis, %)	63–75	75–82

^a Refs. 102, 111, and 113.

^b To convert MPa to psi, multiply by 145.

^c The space velocity is defined as the standard volume of the reactant stream fed per unit time divided by the volume of reactor space filled with catalyst.

^d To convert kPa to mm Hg, multiply by 7.5.

Oxygen-Based Direct Oxidation Process. Even though the fundamental reaction and the ultimate results are the same, there are substantial differences in detail between air- and oxygen-based processes. Virtually all the differences arise from the change in the oxidizing agent from air (ca 20 mol % O₂) to pure oxygen (>95 mol % O₂). Due to the low per pass conversion, the need for complete removal of ethylene oxide by absorption, and the accumulation of nitrogen in the cycle, the air process requires a substantial purge stream. As a direct consequence of this purge stream, the air-based process requires the staged reaction–absorption system described earlier. The oxygen-based process uses substantially pure oxygen, reduces the quantities of inert gases introduced into the cycle, and thereby results in almost complete recycle of the unconverted ethylene (96,106). This eliminates the need for a purge reactor system in an oxygen-based process. However, as in the air-based process, the volume of carbon dioxide formed is about half the volume of ethylene that reacts at a catalyst selectivity of 70–80%. This CO₂ must be eliminated on a continuous basis in order to control its concentration at a fixed acceptable level in the cycle. Concentrations of CO₂ much in excess of 15 mol % adversely influence catalyst activity (96,102). Therefore, in an oxygen-based system, part of the recycle gas leaving the absorber must be treated in a CO₂ removal unit before it is sent back to the main reaction cycle.

In addition to the CO₂ removal unit purge stream, an additional process vent is required to prevent accumulation of argon in the cycle. Argon is a significant impurity in the oxygen supply, and can build up to 30–40 mol % in the cycle gas if no deliberate purge is used (102). When this happens, because of the lower heat capacity of the argon, the cycle gas may enter the flammable region, and as a result the oxygen concentration in the cycle has to be lowered. Consequently, the selectivity of the process is substantially lower (126). In spite of this additional purge, the total vent stream in an oxygen-based process is much smaller than in an air-based unit. The operation of the main reactor can be at much higher ethylene concentration than that possible in

air-based process due to the smaller purge flow. The high ethylene concentration improves the catalyst selectivity because the per pass conversions are lower for a given ethylene oxide production (96,109). The small purge rates in an oxygen-based system operated with very high purity oxygen (99.0–99.5 mol %) make it possible to use cycle diluents of improved heat capacity other than nitrogen (116). These diluents facilitate the use of higher oxygen concentrations in the cycle and, therefore, improve selectivity.

Figure 3 shows a simple schematic diagram of an oxygen-based process. Ethylene, oxygen, and the recycle gas stream are combined before entering the tubular reactors. The basic equipment for the reaction system is identical to that described for the air-based process, with one exception: the purge reactor system is absent and a carbon dioxide removal unit is incorporated. The CO₂ removal scheme illustrated is based on a patent by Shell Oil Co. (127), and minimizes the loss of valuable ethylene in the process.

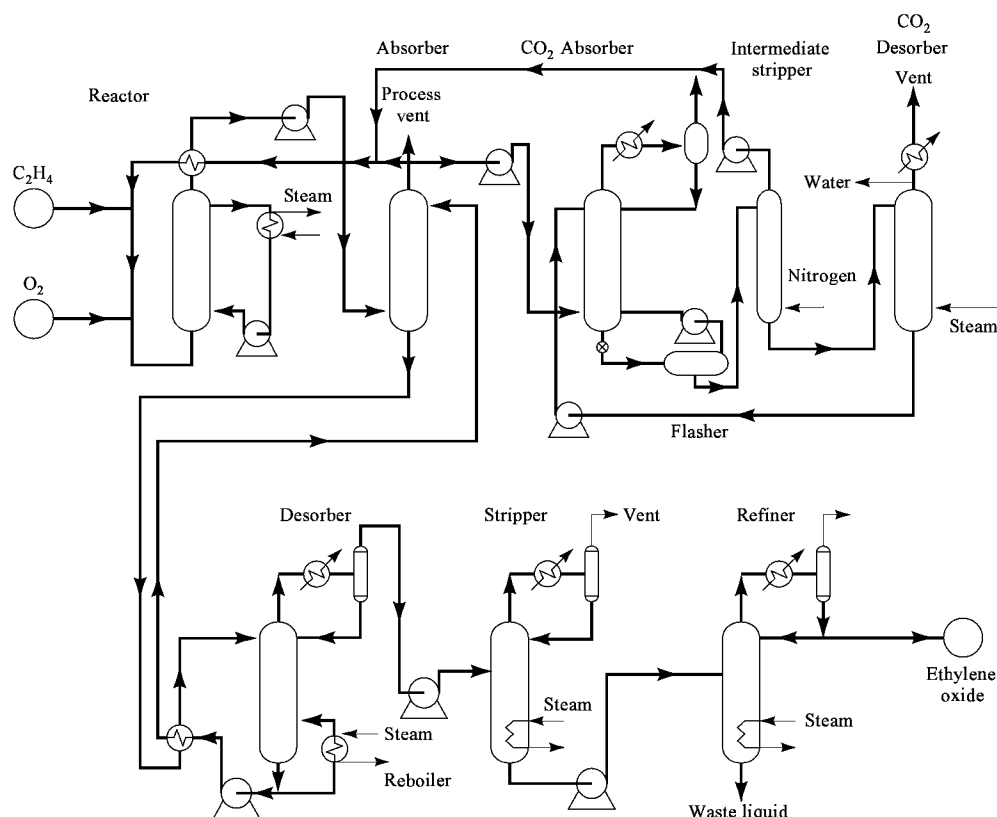


Fig. 3. Oxygen-based direct oxidation process for ethylene oxide (96,102,103,109,117–119,127).

The main process vent stream generally contains a fairly high hydrocarbon concentration, particularly if the diluent is not nitrogen. In such cases, the purge stream can be readily used in a boiler or incinerated in a furnace without supplemental fuel (106,128). The ethylene oxide recovery and refining sections for both the air- and oxygen-based processes are almost identical. As with the air-based process, specific operating conditions for the reaction system are proprietary; however, the general ranges reported in the literature and patents are summarized in Table 10.

Process Technology Considerations. Innumerable complex and interacting factors ultimately determine the success or failure of a given ethylene oxide process. Those aspects of process technology that are common to both the air- and oxygen-based systems are reviewed below, along with some of the primary differences.

Ethylene Oxide Catalysts. Of all the factors that influence the utility of the direct oxidation process for ethylene oxide, the catalyst used is of the greatest importance. It is for this reason that catalyst preparation and research have been considerable since the reaction was discovered. There are four basic components in commercial ethylene oxide catalysts: the active catalyst metal; the bulk support; catalyst promoters that increase selectivity and or activity and improve catalyst life; and inhibitors or antocatalysts that suppress the formation of carbon dioxide and water without appreciably reducing the rate of formation of ethylene oxide (105).

Silver-containing catalysts are used exclusively in all commercial ethylene oxide units, although the catalyst composition may vary considerably (129). Nonsilver-based catalysts such as platinum, palladium, chromium, nickel, cobalt, copper kenenide, gold, thorium, and antimony have been investigated, but are only of academic interest (98,130–135). Catalysts using any of the above metals either have very poor selectivities for ethylene oxide production at the conversion levels required for commercial operation, or combust ethylene completely at useful operating temperatures.

A variety of different procedures have been reported for silver catalyst preparation on relatively inert support materials. The different methods are (1) precipitation of silver oxide from aqueous silver nitrate or other salt solutions by alkali or alkaline-earth compounds (136,137); (2) thermal decomposition of silver salts, in particular silver oxalate, silver carbonate, or organic salts of silver (137,138); (3) reduction of silver salts by hydrogen, formaldehyde, hydrazine, or hydroxylamine (139,140); (4) electrolysis of silver salt solutions (141); and (5) selective removal of secondary metals from silver containing alloys (142,143). The silver is added to the support either as a coating from a suspension or by impregnation with a solution. The coating procedure is claimed to give a catalyst of higher silver content and initial activity. However, the catalyst is susceptible to silver loss by abrasion and tends to lose selectivity to ethylene oxide after use for several months (144). Impregnation of the support with a silver salt combined with an organic reducing agent appears to be more popular in the patent literature. Some investigators have used a combination of coating and impregnation procedures (144–146).

The chemical and physical properties of the support strongly dictate the performance of the finished catalyst. Although nonsupported silver catalysts have been advocated in some patents (147), it is unlikely that they are used commercially since pure silver tends to sinter at reaction temperatures with a resultant activity loss (102). For commercial operation, the preferred supports are alundum (α-alumina) and silicon carbide (129,131). Other supports are glass wool, quartz, carborundum, and ion-exchange zeolites (see MOLECULAR SIEVES). The surface area, porosity, and pore size of the support influence the size of the silver particles on the support and, therefore, affect the performance of the final catalyst (148). High surface area supports (3–100 m²/g) generally yield poor ethylene oxide catalysts presumably because ethylene reacts in the pores from which the ethylene oxide is released slowly. The combination of slow product release and poor heat conductivity of high surface area supports is claimed to result in the combustion of ethylene oxide (149).

Silver alone on a support does not give rise to a good catalyst (150). However, addition of minor amounts of promoter enhance the activity and the selectivity of the catalyst, and improve its long-term stability. Excess addition lowers the catalyst performance (151,152). Promoter formulations have been studied extensively in the chemical industry. The most commonly used promoters are alkaline-earth metals, such as calcium or barium, and alkali metals such as cesium, rubidium, or potassium (153). Using these metals in conjunction with various counter anions, selectivities as high as 82–87% were reported. Precise information on commercial catalyst promoter formulations is proprietary (154–156).

Many organic compounds, especially the halides, are very effective for suppressing the undesirable oxidation of ethylene to carbon dioxide and water, although not significantly altering the main reaction to ethylene oxide (150). These compounds, referred to as catalyst inhibitors, can be used either in the vapor phase during the process operation or incorporated into the catalyst manufacturing step (102). The inhibitor plays a significant role in the process, although the mechanism of inhibitor action is still a subject of debate (157). Important gas-phase inhibitors are ethylene dichloride, ethylene dibromide, other alkyl halides, aromatic hydrocarbons, amines, and organometallic compounds (150,158). In a study of the effect of ethylene dichloride on catalyst activity, it was found that small amounts improved catalyst performance; however, excess amounts of ethylene dichloride deactivated the catalyst (129).

In order to assure control of the reaction, the vapor-phase inhibitor concentration must be closely controlled in the ppm range. Although several compounds have been claimed to be useful, it is likely that commercial processes use only ethylene dichloride or some of the simpler chlorinated aromatics (102). In general, the choice between inhibitors is not based on their differences in performance, but rather on the designers preference for dealing with the type of control problems each inhibitor system imposes (102).

Temperature and Thermal Factors. Temperature is used to control two related aspects of the reaction: heat removal from the reactor bed and catalyst operating temperature. The reactor temperature is controlled through the use of a heat-transfer fluid on the reactor shell. The coolant used in most recent designs is boiling water. Boiling water provides good heat transfer and improved safety over previous reactor designs using either boiling or circulating organic heat-transfer fluids. Control of the catalyst operating temperature is necessary to prevent catalyst damage such as sintering or tube damage resulting from excessive temperatures in the catalyst bed. Localized hot spots of 100–300°C above the coolant temperature can form in the catalyst bed without adequate temperature control.

Heat removal from the reactor is necessary for stable operation. Heat is removed through two processes: sensible heat of the process gas and heat transfer to the coolant. Stable operation of the reactor requires balancing these removal processes with the heat generated in the reaction. The undesired complete oxidation of ethylene is the primary source of heat, with a heat of reaction of 1419.0 kJ/mol of ethylene compared with the partial oxidation heat of reaction of 105.2 kJ/mol (25 kcal/mol). Since the activation energy for complete combustion (16.8 kJ/mol) is greater than for partial oxidation (15.2 kJ/mol), and higher temperatures increase the rate of ethylene oxide combustion, the catalyst selectivity drops as temperature increases (159). This compounds the increased heat generation with increasing catalyst temperature. For example, 25 t/h of ethylene oxide production at 75% selectivity generates 91.3 MW; increasing the temperature to increase production to 30 t/h will decrease the catalyst selectivity, say to 70%, and the heat generated (135 MW) for this 20% increase in production is a 48% increase. Therefore, stable operation of the ethylene oxide reactor requires stability in the coolant temperature control.

The reactor stability is also affected by the internal temperature profile in the tubular catalyst bed. The highest reaction rate occurs near the inlet, where the partial pressure of the components is the highest (160). This results in the highest heat generation rate and therefore the maximum (bed temperature–coolant) temperature. Stable operation of the reactor requires keeping this peak temperature difference to less than 30–40°C. Several techniques have been proposed for resolving this problem. For example, a gradation in the catalyst activity down the length of the reactor could result in a uniform temperature profile and improved catalyst selectivity (161). Other solutions propose improving the reactor thermal conductivity (160).

Space Velocity. Space velocity has a strong effect on the process economics. It establishes the reactor size and pressure drop, affecting compression costs. The optimum space velocity is a function of energy costs, reaction rate, and selectivity. As space velocity increases, the reactor heat-transfer coefficient, selectivity, and compression costs increase while conversion decreases. The optimization of space velocity is highly integrated into the design of the unit since all of the process variables, pressure, temperature, catalyst, etc, affect the optimum. The literature claims benefits in operating in the range of 1000–4500 h^{–1} (110,153,162–165).

Operating Pressure. Operating pressure has a marginal effect on the economics of the ethylene oxide process. High pressure increases production due to higher gas density, increases heat transfer, increases ethylene oxide and carbon dioxide recovery in the absorbers, and lowers compression costs. Also, since the total number of moles decreases in the formation of ethylene oxide from ethylene and oxygen, high pressure is consistent with high conversion. However, high pressures reduce the flammable limit of the process gas as well as increase equipment costs. Typical commercial pressures are 1–2 MPa (10–20 atm) (110,158,165–167).

Cycle Diluents. Air process technology uses nitrogen as the diluent gas. The amount of nitrogen that enters the process in the air feed cannot be economically diluted (97).

The choice of diluent for the oxygen process is based on the thermal properties of the gas. The small process purge makes it economically possible for the process to operate under a wide variety of ballast gases. Several gases have been proposed in the patent literature, including methane (168) and ethane (169). Both of these gases have the advantage of higher specific heat and thermal conductivity than nitrogen, raising the flammable limit and reducing the peak temperature difference in the catalyst bed. However, these diluents lead to less steam generation from the reactor coolant due to more sensible heat loss from the reactor, and increase the risk of forming a flammable vapor cloud following a rupture or leak. They also require additional purification systems to prevent sulfur and other contaminants from poisoning the catalyst (170). Finally, higher ethane concentrations will increase the amount of organic chlorides that are needed. This can result in material of construction and product quality problems, as well as adversely affecting catalyst life.

Raw Material Purity Requirements. The oxygen process has four main raw materials: ethylene, oxygen, organic chloride inhibitor, and cycle diluent. The purity requirements are established to protect the catalyst from damage due to poisons or thermal runaway, and to prevent the accumulation of undesirable components in the recycle gas. The latter can lead to increased cycle purging, and consequently higher ethylene losses.

Typical ethylene specifications call for a minimum of 99.85 mol% ethylene. The primary impurities are usually ethane and methane. A methane limit is largely unnecessary; however, care should be taken to restrict the amount of ethane since high ethane concentration will lead to increased chloride inhibitor concentration, which adversely affects product quality, catalyst life, and materials of construction. Impurities that strongly affect catalyst performance and reactor stability include acetylene, propylene, hydrogen, and sulfur. Acetylene causes catalyst coking at very low concentrations (158). Carbonaceous deposits can also be caused by heavy hydrocarbons if present. Propylene is more reactive than ethylene and will oxidize to a wide range of products, including aldehydes that lower product quality (171,172). Hydrogen and carbon monoxide can lead to hot-spotting of the catalyst, and sulfur is a nonreversible poison for silver-based catalyst (120).

Oxygen specifications can vary depending on the economics of the process. The dominant impurity in oxygen is argon. As the argon concentration in the reactor feed increases, the flammable limit of the gas decreases. Therefore, the cycle gas purge flow rate is established to maintain a given concentration of argon in the recycle stream. The optimization of oxygen purity balances the increased ethylene losses in the cycle gas purge and reduced oxygen concentration in the reactor feed gas against the cost of higher purity oxygen (96). Typical oxygen purity is 95–99.95 mol%. However, several patents have proposed using membranes and adsorption technologies to recover the ethylene in the cycle purge gas and recycle it to the process (173–175). In this case, the economics of the process would lead to reduced oxygen purities.

Organic chloride and cycle diluent specifications are less critical since the flows are significantly less. The organic chloride specifications must prevent gross contamination as well as the potential of solids that would lead to plugging. The cycle diluent must also be free of gross contamination as well as significant catalyst poisons such as sulfur (170).

The air process has similar purity requirements to the oxygen process. The ethane content of ethylene is no longer a concern, due to the high cycle purge flow rate. Air purification schemes have been used to remove potential catalyst poisons or other unwanted impurities in the feed.

Ethylene Oxide Recovery. An economic recovery scheme for a gas stream that contains less than 3 mol% ethylene oxide (EO) must be designed. It is necessary to achieve nearly complete removal since any ethylene oxide recycled to the reactor would be combusted or poison the carbon dioxide removal solution. Commercial designs use a water absorber followed by vacuum or low pressure stripping of EO to minimize oxide hydrolysis. Several patents have proposed improvements to the basic recovery scheme (176–189). Other references describe how to improve the scrubbing efficiency of water or propose alternative solvents (180,181).

Ethylene Oxide Purification. The main impurities in ethylene oxide are water, carbon dioxide, and both acetaldehyde and formaldehyde. Water and carbon dioxide are removed by distillation in columns containing only rectifying or stripping sections. Aldehydes are separated from ethylene

oxide in large distillation columns. The size of the column is related to the high degree of separation required for meeting the product quality demand of poly(ethylene glycol) and detergent ethoxylate manufacturers, typically <30 ppm aldehydes. Refining ethylene oxide to meet this specification increases the amount of formaldehyde in the final product relative to acetaldehyde due to its higher volatility (182). In addition to modifications of conventional distillation, other technologies for removing aldehyde from ethylene oxide have been proposed, including molecular sieves and extractive distillation (183–186).

Process Safety Considerations. Unit optimization studies combined with dynamic simulations of the process may identify operating conditions that are unsafe regarding fire safety, equipment damage potential, and operating sensitivity. Several instances of fires and deflagrations in ethylene oxide production units have been reported in the past (160). These incidents have occurred in both the reaction cycle and ethylene oxide refining areas. Therefore, ethylene oxide units should always be designed to prevent the formation of explosive gas mixtures.

The safe operating ranges of the unit are dependent on all of the process parameters: temperature, pressure, residence time, gas composition, unit dynamic responses, instrumentation system, and the presence of ignition sources (160). The ethylene oxide reaction cycle operates close to the flammable limit. Higher oxygen concentrations yield higher activity and efficiency but more closely approach the flammable limit. One of the more sensitive areas of the unit's design is oxygen mixing. The oxygen concentration must pass through the flammable region while diluting down to the operating level. This is particularly acute in the oxygen process. Therefore, the oxygen mixing step is highly instrumented, and requires a fairly complicated control scheme to assure safe mixing. Another sensitive area is the final refining of high purity ethylene oxide. As with any reactive chemical, it is prudent to use the lowest temperature heat source practical when refining ethylene oxide. Proper operation of the reboiler, including control of liquid level, is critical. The presence of specific forms of iron oxide in contact with ethylene oxide vapor can lead to highly exothermic reactions that can initiate the explosive decomposition of ethylene oxide. Proper selection of insulation is critical. Ethylene oxide leaks in porous insulation can react exothermically with water in the insulation, and the resulting glycol may spontaneously ignite at temperatures greater than 60°C (190,191). The selection of the safe operating conditions and design of effective process safety systems is a complex task that requires extensive laboratory testing to determine the effect of the various process parameters on explosibility as well as proven commercial experience.

Environmental Considerations. A detailed study of the environmental considerations in the manufacture of ethylene oxide by the direct oxidation of processes is described in Reference 108. The primary air emissions from the formation of ethylene oxide by direct oxidation are ethylene, ethylene oxide, carbon dioxide, and ethane. Traces of NO_x and SO_x from pollution control and process machinery operations have also been reported. The largest source of organic emissions to the atmosphere in either the oxygen or air process is the cycle purge. In the oxygen process, this is a low volume, high hydrocarbon stream that can be readily used as boiler fuel since it contains 80–85 mol % hydrocarbons (97). The air process cycle purge is the vent from the last purge reactor. Relative to the oxygen process cycle purge, this is a high volume stream that contains low concentrations of hydrocarbons (<2 mol %). A catalytic converter may be added to this stream to reduce the organic emissions.

Other air emissions occur in the ethylene oxide recovery and purification sections. In the purification section, a light end vent is taken from the ethylene oxide distillation train. This vent contains mostly carbon dioxide with some ethylene and ethylene oxide. In some units this stream is passed through a vent scrubber to remove ethylene oxide, recompressed, and fed back to the reaction cycle to recover the ethylene. The other emission source is the carbon dioxide regenerator vent in the oxygen process. This vent includes ethylene and ethylene oxide that are absorbed from the process gas. Some designs call for the installation of a single-stage flash between the CO₂ absorber and regenerator to reduce the ethylene emissions.

Relative to the process streams, emissions from auxiliary equipment and flares are small. Some ethylene oxide units use gas-fired turbines to feed air or ethylene (109). These result in unburned hydrocarbon and possible NO_x emissions. Also, most ethylene oxide units have flares to vent the process gas during upsets. Data is scarce, but estimates indicate that flaring of process gas occurs once to twice a year (109).

Liquid emissions from ethylene oxide units originate in the recovery section. The water of reaction from complete combustion of ethylene must be purged from the oxide absorber water cycle. This stream contains glycol, organic salts, aldehydes, and ethylene oxide. The location of the purge stream is selected to minimize ethylene oxide and glycol emissions. This stream is readily biodegradable (109). Direct oxidation processes that operate at lower temperatures generate fewer impurities, and therefore have lower organic loads on the process waste treatment unit. Several technologies have been proposed to reduce the amount of organics in the waste, either by distillation or the use of membranes to recover the contained glycol (187–189,192–194).

Air vs Oxygen Process Differences and Economics. The relative economics of the air vs oxygen process are reported (97). Two process characteristics dictate the difference in the capital costs for the two processes. The air process requires additional investment for the purge reactors and their associated absorbers, and for energy recovery from the vent gas. However, this is offset by the need for an oxygen production facility and a carbon dioxide removal system for an oxygen-based unit. In a comparison of necessary investments for medium to large capacity units (>20,000 t yr), oxygen-based plants have a lower capital cost even if the air-separation facility is included (97). However, for small- to medium-scale plants, the air process investment is smaller than that required for the oxygen process and the air-separation unit, unless the oxygen is purchased from a large air-separation unit serving many customers.

There are also operating cost considerations that differ significantly among the two processes. The costs of silver catalyst, oxygen, and ethylene are critical factors determining the relative economics. For a given catalyst type, the oxygen process operates at a higher selectivity and requires a smaller volume of catalyst. Even though the cost of ethylene comprises ca 60% of the total manufacturing cost in both processes, the incremental product between the air and oxygen processes is influenced only slightly by changes in the price of ethylene. For example, in 1976 it was estimated that an ethylene price increase of 2.20 ¢ kg raised the product cost for air oxidation by only 0.088 ¢ kg in excess of that of an oxygen-based unit (96). On the other hand, the price of oxygen has a much more significant effect on the economics of an oxygen unit. A change in the oxygen price by 0.22 ¢ kg altered the product cost by 0.243 ¢ kg.

The oxygen-process plant has no high pressure purge gas stream of sufficient volume to make energy recovery attractive as in an air process. The oxygen process also has a considerable steam requirement for carbon dioxide stripping in the CO₂ removal unit. The total compression costs for an oxygen-based process, including the air-separation unit, are slightly higher than for an air-based system (160).

Purity of the feedstocks (oxygen and ethylene) also determine the relative economics of the air- and oxygen-based processes. If the oxygen purity is low, the volume of the ethylene-rich purge gas is increased markedly, and the oxygen-based process becomes unattractive. In addition to requiring a high oxygen purity, the ratio of argon to argon plus nitrogen in the oxygen feed is critical to the attainment of high yields in the oxygen process (195). Lower values of the ratio improve yields by ca 1–2%. For large-scale ethylene oxide processes, yield changes of even a fraction of a percent can have an impact on the overall process economics. In general, as the ethylene purity decreases, the air process becomes more attractive (97). However, an air-based process may require an air purification system if contaminants such as sulfur, halogens, and heavier hydrocarbons are present.

Both air and oxygen processes can be designed to be comparable in the following areas: product quality, process flexibility for operation at reduced rates, and on-stream reliability (97,182). For both processes, an on-stream value of 8000 h yr is typical (196). The reliability of the oxygen-based system is closely linked to the reliability of the air-separation plant, and in the air process, operation of the multistage air compressor and power recovery from the vent gas is crucial (97).

For the same production capacity, the oxygen-based process requires fewer reactors, all of which operate in parallel and are exposed to reaction gas of the same composition. However, the use of purge reactors in series for an air-based process in conjunction with the associated energy recovery system increases the overall complexity of the unit. Given the same degree of automation, the operation of an oxygen-based unit is simpler and easier if the air-separation plant is outside the battery limits of the ethylene oxide process (97).

From the preceding discussion, it is clear that no meaningful generalizations can be made regarding the overall superiority of either the air- or oxygen-based process.

Other Processes.

Chlorohydrin Process. Ethylene oxide is produced from ethylene chlorohydrin by dehydrochlorination using either sodium or calcium hydroxide (160). The by-products include calcium chloride, dichloroethane, bis(2-chloroethyl) ether, and acetaldehyde. Although the chlorohydrin process appears simpler, its capital costs are higher, largely due to material of construction considerations (197).

Arsenic-Catalyzed Liquid-Phase Process. An arsenic catalyst liquid-phase process for olefin oxides has been patented by Union Carbide (198). The selective epoxidation of ethylene by hydrogen peroxide in a 1,4-dioxane solvent in the presence of an arsenic catalyst is claimed. No solvent degradation is observed. Ethylene oxide is the only significant product detected. The catalyst used may be either elemental arsenic, an arsenic compound, or both.

Thallium-Catalyzed Epoxidation Process. The use of Tl(III) for olefin oxidation to yield glycols, carbonyls, or epoxides is well known (199). Because the epoxidation with Tl(III) is stoichiometric to produce Tl(I), reoxidation is needed. Halcon has patented processes based on such epoxidation to yield ethylene oxide (200–203). The primary benefits of such a process are claimed to be high yields of ethylene oxide, flexibility to produce either propylene oxide or ethylene oxide, and the potential of a useful by-product (acetaldehyde). Advances using organic hydroperoxides in place of oxygen for reoxidation offer considerable promise, since reaction rates are rapid and low pressures can be used.

Lummus Hypochlorite Process. A Lummus patent claims a process for propylene oxide or ethylene oxide using *tert*-butyl hypochlorite (204). The chemistry for this new process parallels the classical chlorohydrin technology with brine recycle. Advantages claimed are high ethylene yield, reduced reactor size, and lower steam requirements for the saponification step. However, disadvantages include high capital cost for a large chlorine plant, difficult and energy-intensive distillation steps, and losses of *tert*-butyl alcohol in the process.

Liquid-Phase Epoxidation with Hydroperoxides. Molybdenum, vanadium, and tungsten have been proposed as liquid-phase catalysts for the oxidation of the ethylene by hydroperoxides to ethylene oxide (205). *tert*-Butyl hydroperoxide is the preferred oxidant. The process is similar to the arsenic-catalyzed route, and includes the use of organometallic complexes.

Electrochemical Process. Several patents claim that ethylene oxide is produced in good yields in addition to faradic quantities of substantially pure hydrogen when water and ethylene react in an electrochemical cell to form ethylene oxide and hydrogen (206–208). The only raw materials that are utilized in the ethylene oxide formation are ethylene, water, and electrical energy. The electrolyte is regenerated *in situ*, ie, within the electrolytic cell. The addition of oxygen to the ethylene is activated by a catalyst such as elemental silver or its compounds at the anode or its vicinity (206). The common electrolytes used are water-soluble alkali metal phosphates, borates, sulfates, or chromates at ca 22–25°C (207). The process can be either batch or continuous (see ELECTROCHEMICAL PROCESSING).

Unsteady-State Direct Oxidation Process. Periodic interruption of the feeds can be used to reduce the sharp temperature gradients associated with the conventional oxidation of ethylene over a silver catalyst (209). Steady and periodic operation of a packed-bed reactor has been investigated for the production of ethylene oxide (210). By periodically varying the inlet feed concentration of ethylene or oxygen, or both, considerable improvements in the selectivity to ethylene oxide were claimed.

Fluid-Bed Direct Oxidation Process. The use of a fluidized bed for ethylene oxide production was reported to produce good temperature control and inhibition of side reactions using a granular silver catalyst (211). Several additional fluid-bed processes that claim improved product yields were patented later (212–215). However, only one process, developed by Vulcan Atlantic, was reported to have been successfully demonstrated in pilot-scale equipment (216). The pilot fluid-bed reactor is described as a multitubed converter that produces uniform heat transfer to a fluid circulating in the shell of the converter, and minimizes back-mixing of the fluidized catalyst. A novel circulating fluid-bed process and the associated reaction apparatus have been patented (217–219). By employing high gas velocities, the ethylene oxide productivity per unit catalyst volume is claimed to be three to four times greater than the maximum reported number for conventional fixed-bed tubular reactor (219,220). In spite of the recent rapid advances in fluidization technology, no commercial fluid-bed ethylene oxide processes have been reported.

Biological. Several recent patents have claimed the production of ethylene oxide from a wide variety of raw materials using enzymatic catalysts (221–224). However, no commercial production routes based on biological mechanisms have been proposed.

Economic Aspects

United States production of ethylene oxide in 1990 was 2.86×10^6 metric tons. Approximately 16% of the United States' ethylene (qv) production is consumed in ethylene oxidation, making ethylene oxide the second largest derivative of ethylene, surpassed only by polyethylene (see OLEFIN POLYMERS). World ethylene oxide capacity is estimated by country in Table 11. Total world capacity in 1992 was ca 9.6×10^6 metric tons.

Table 11. World Production Capacities for Ethylene Oxide^a

Country	Total capacity, 10 ³ t yr	World capacity, %	Number of companies ^c	Capacity by process, ^b 10 ³ t yr			
				UCC&P	Shell	Scientific Design	Other
United States	3452	35.8	12	1052	1323	807	
Canada	475	4.9	2	290			
Mexico	303	3.1	5			303	
Brazil	149	1.5	1			149	
Belgium	120	1.1	2	120			
France	170	1.8	1		170		
Germany	835	8.7	5		745	90	
Italy	60	0.6	2	20			40
Netherlands	340	3.5	2		190		
United Kingdom	230	2.4	1		230		
Sweden	40	0.4	1			40	
Spain	100	1.0	1		100		
Bulgaria	85	0.9	S			75	10
Czechoslovakia	55	0.6	S		35		20
Poland	80	0.8	S		55		25
Romania	70	0.7	S			70	
Commonwealth of Independent States	435	4.5	S			380	55
Japan	759	7.9	4		449	100	210
People's Republic of China	385	4.0	S			325	60
Taiwan	190	2.0	2	125	35	30	
India	178	1.8	4	45	91	42	
North Korea	10	0.1	S				10
South Korea	340	3.5	3		180	160	
Australia	30	0.3	1			30	
Singapore	80	0.8	1		80		
Turkey	50	0.5	1		50		
Saudi Arabia	630	6.5	2		306	270	
Total	9651	99.7		1652	4093	2871	430

^a Capacities are estimated as of Dec. 1992.
^b Dow process capacity is 270,000, 185,000, and 150,000 in the United States, Canada, and the Netherlands, respectively.

^c S all capacity is government owned.

More than 98% of the total production of ethylene oxide is converted to derivatives. The consumption pattern of ethylene oxide in the United States is shown in Table 12. About 60–65% of the ethylene oxide is consumed in manufacture of ethylene glycols (99,225). The largest markets for ethylene glycol are antifreeze (qv) and polyester (qv) intermediates. Each of these markets account for approximately 27% of the United States' ethylene oxide consumption. Other derivatives of ethylene oxide include glycol ethers, ethanalamines, and surfactants. Principal uses of these derivatives are shown in Table 12, as well as the expected U.S. annual growth rates through 1995 of each derivative. Antifreeze is predicted to have zero growth. The fastest growing market is expected to be ethylene glycol for polyesters (5% annually). Overall growth in U.S. demand for ethylene oxide is ca 3% yr through 1995 (99). Worldwide ethylene oxide demand is expected to grow at a slightly higher 5% rate, highlighted by a 10% annual growth in ethylene glycol for Far East polyester markets.

Table 12. United States Consumption Pattern for Ethylene Oxide^a

Use	Consumption, %	Projected annual growth rate, % ^b	Principal use
ethylene glycol for antifreeze	27	0	automotive coolant
ethylene glycol for polyester	27	5	polyester fibers, films, and containers
glycol ethers	6	0	solvents for lacquers, resins, enamels, epoxy coatings, water-based coatings, var-nishes, and printing inks; antiicing agent in jet fuel
ethanolamines	9	3	soaps and detergents; gas conditioning
surfactants	14	3	detergents
all other	17	3	

^a Consumption pattern estimated for 1990, Ref. 226.

^b 1990–1995.

The U.S. price history for merchant ethylene oxide sales is shown in Figure 4. It can be seen that ethylene oxide pricing generally reflects the cost of ethylene and, of course, supply–demand forces. In the late 1980s, several U.S. producers announced their withdrawal from the merchant ethylene oxide market.

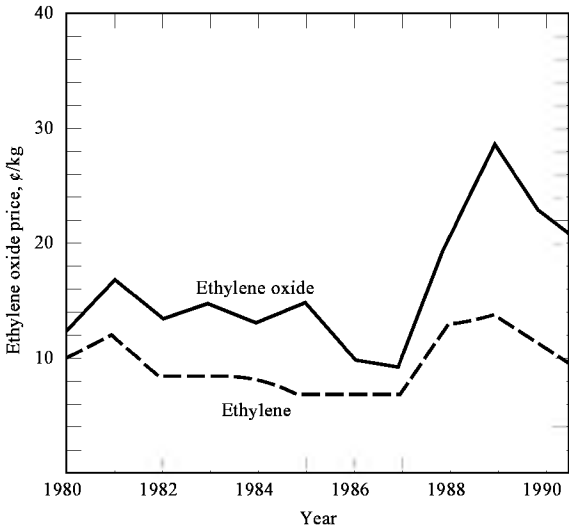


Fig. 4. Price history for ethylene oxide sales.

Shipment and Storage

Small shipments of ethylene oxide are made in either compressed gas cylinders up to ~0.1 m³ (30 gal) or in 1A1 steel drums (61 gal). Very large shipments >40 m³ (10,000–25,000 gal) are made in insulated, type 105J100W or other DOT approved tank cars. For further information on the shipping and handling of ethylene oxide, see References 9 and 227.

Storage. Carbon steel and stainless steel should be used for all equipment in ethylene oxide service. Ethylene oxide attacks most organic materials (including plastics, coatings, and elastomers); however, certain fluoroplastics are resistant and can be used in gaskets and O-rings. See Reference 9 for a list of materials that are compatible with ethylene oxide.

Storage tanks should be designed in accordance with the ASME code for unfired pressure vessels. All-welded construction is recommended. Ethylene oxide storage tanks should be electrically grounded, isolated from potential fire hazards, and equipped with pressure relief devices. New equipment should be cleaned of iron oxide and immediately purged with inert gas.

Ethylene oxide storage tanks are pressurized with inert gas to keep the vapor space in a nonexplosive region and prevent the potential for decomposition of the ethylene oxide vapor. The total pressure that should be maintained in a storage tank increases with liquid temperature, since the partial pressure of ethylene oxide will also increase. Figure 5 shows the recommended minimum storage pressures for liquid ethylene oxide under nitrogen or methane blanketing gas.

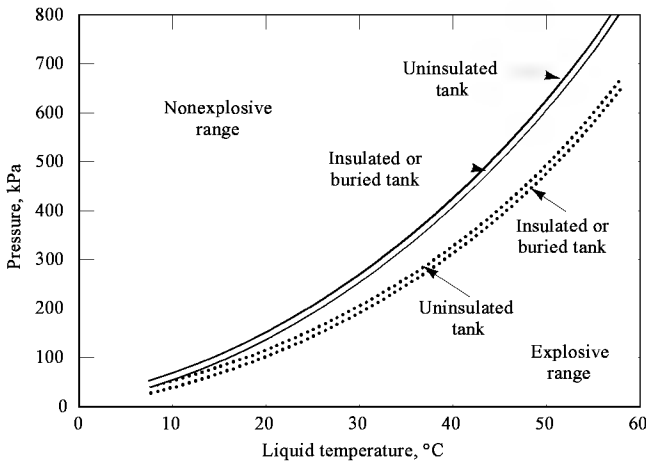


Fig. 5. Recommended safe storage pressures for liquid ethylene oxide under nitrogen (—) or methane (···) blanketing (9). To convert kPa to psi, multiply by 0.145.

Polymerization of ethylene oxide can occur during storage, especially at elevated temperatures. Contamination with water, alkalies, acids, amines, metal oxides, or Lewis acids (such as ferric chloride and aluminum chloride) can lead to runaway polymerization reactions with a potential for failure of the storage vessel. Therefore, prolonged storage at high temperatures or contact with these chemicals must be avoided (9).

Worldwide capacity for ethylene oxide should be boosted by $ca\ 2 \times 10^6\ t\ yr$ to a total of $11.6 \times 10^6\ t\ yr$ by 1995. Projects underway are listed in Table 13. Additions have been announced by Union Carbide in the United States and Canada, Formosa Plastics and Shell in the United States, BASF in Belgium (rebuild of an existing unit), and SABIC in Saudi Arabia (100,101). Expansions have also been announced in South Korea and India, and may occur in other countries as well.

Table 13. Planned Ethylene Oxide Capacity^a

Company	Planned site	Capacity, 10 ³ t yr	Technology
UCC&P	Taft, La.	190	UCC&P
UCC&P	Prentiss, Canada	218	UCC&P
Formosa Plastics	Pt. Comfort, Tex.	240	Scientific Design
BASF	Antwerp, Belgium	250	Shell
SABIC	Al-Jabail, Saudi Arabia	250	Shell
Shell	Geismar, La.	204	Shell

^a Ref. 100.

Specifications, Analytical, and Test Methods

Ethylene oxide is sold as a high purity chemical, with typical specifications shown in Table 14. This purity is so high that only impurities are specified. There is normally no assay specification. Proper sampling techniques are critical to avoid personal exposure and prevent contamination of the sample with trace levels of water. A complete review and description of analytical methods for pure ethylene oxide is given in Reference 228.

Table 14. Specifications for Ethylene Oxide^a

Property	Value
acidity	0.002 wt % max, calculated as acetic acid
aldehydes	0.003 wt % max, calculated as acetaldehyde
water	0.03 wt % max
residue	0.005 g / 100 mL max
color	10 Pt–Co max
suspended matter	substantially free

^a Ref. 9.

The near-ir spectrum of ethylene oxide shows two peaks between 1600–1700 nm, which are characteristic of an epoxide. Near-ir analyzers have been used for verification of ethylene oxide in railcars. Photoionization detectors are used for the determination of ethylene oxide in air (229–232). These analyzers are extremely sensitive (lower limits of detection are ~0.1 ppm) and can compute 8-h time-weighted averages (TWA₈).

Health and Safety Factors

Toxicology. An excellent review of the toxicity and health assessment of ethylene oxide has been compiled (233). Ethylene oxide (EO) can be relatively toxic as both a liquid and gas. Inhalation of ethylene oxide in high concentrations may be fatal. Estimates of lethal ethylene oxide inhalation levels in animals depend on the duration of exposure. The reported 4-h LC₅₀ values for rats, mice, and dogs are 1460, 835, and 960 ppm, respectively (234). More recent information (235) indicates that the 1-h LC₅₀ in rats is approximately 5000 ppm.

Inhalation exposure to high concentrations of ethylene oxide has been reported to result in respiratory system irritation and edema (236). Depending on the degree of exposure, there may be stinging of the nose and throat, coughing, and chest tightness. Also, exposure may cause lung injury and delayed onset of pulmonary edema. In long-term studies of animals exposed to less than 100 ppm of ethylene oxide and in human studies, no evidence of injury has been reported for the cardiovascular system, liver, or kidney. There is some evidence that occupational exposure to high levels of ethylene oxide can result in cataracts (237). Neurological effects have also been reported in association with recurrent human (238) and animal (239) inhalation exposures to ethylene oxide. Again, depending on the degree of exposure, headache, nausea, vomiting, diarrhea, dizziness, loss of coordination, convulsion, or coma may occur. The onset of illness is rapid in severe exposures, but may be delayed after moderate exposure. In the reports of human peripheral

neurotoxic effects or central nervous system toxicity, most cases have shown a marked improvement on removal from further exposure.

Ethylene oxide has been shown to produce mutagenic and cytogenic effects in a variety of test systems (226). An increased frequency of chromosomal aberrations in peripheral lymphocytes of monkey exposed to ethylene oxide for 104 weeks has been reported (240). In mice, it is an effective inducer of chromosome breaks leading to dominant-lethal mutations. In addition, ethylene oxide has been shown to induce heritable effects in the heritable translocation test conducted in mice exposed to ethylene oxide by inhalation (241,242). In this study, male mice were exposed to ethylene oxide ranging from 165 to 300 ppm for 6 h per day 5 or 7 days week for 8.5 weeks. Ethylene oxide has also been shown to bind to proteins (243) as well as to DNA (244). Several studies on ethylene oxide-exposed workers have demonstrated an increased incidence of chromosomal aberrations and sister chromatid exchanges; the relevance of such effects to human health evaluation is currently uncertain.

In an animal study of rats exposed by inhalation to ethylene oxide at 10, 33, or 100 ppm for approximately two years (245), and in a separate chronic rat study in which rats were exposed to 50 or 100 ppm of ethylene oxide (240), increased incidences of mononuclear cell leukemia, peritoneal mesothelioma, and various brain tumors have been reported. In an NTP (246) two-year inhalation study of mice at 50 and 100 ppm, alveolar bronchiolar carcinomas and adenomas, papillary cystadenomas of the harderian gland, and malignant lymphomas, uterine adenocarcinomas, and mammary gland tumors were increased in one or both exposure groups.

When the data as a whole are reviewed for studies on humans exposed to ethylene oxide, no conclusion can be made that there is an increase in mortality associated with those exposed to ethylene oxide. Two Swedish studies (247,248) indicated an increase in leukemia for workers exposed to multiple chemicals including ethylene oxide; however, in a recent larger Swedish study (249) of workers exposed to only ethylene oxide, there was no association of any type of cancer increase for these workers. In a recent study sponsored by NIOSH, there was no significant increase in mortality observed for cancer when all types are combined or for certain individual types of cancer, even for those people who worked the longest and were observed the longest. However, a statistically significant increase in mortality from certain types of lymphoma was observed for male workers. This is contrary to the results observed for female workers. In addition, four other cohort studies of ethylene oxide-exposed workers have been published (250–253), but no unequivocal increase in the risk of cancer was observed.

Developmental toxicity inhalation tests have been conducted using rats or rabbits (254,255). No teratogenic effects were observed. The only developmental effect noted was decreased fetal weights and delayed ossification. In a one-generation inhalation study (256) at the highest concentration tested (100 ppm), there were signs of embryo- or fetotoxicity, as demonstrated by a decrease in the number of implantation sites and a decrease in litter size. Dominant lethal mutagenic effects in rodents have also been reported at high vapor concentrations (257,258). Testicular atrophy in guinea pigs and slight degeneration of seminiferous tubules in rats exposed to high vapor concentrations of ethylene oxide has been observed (239). Data in humans are limited. In an epidemiologic study (259), the spontaneous abortion rates in ethylene oxide sterilizer hospital personnel were reported to be increased; however, because of a number of problems with this study, both in design and analysis, the results are difficult to interpret at best (233).

Dermal exposure information has been collected from case reports of industrial accidents. Concentrated ethylene oxide evaporates rapidly from the skin and produces a freezing effect, often compared to frostbite, leaving burns ranging from first to third degree severity (260,261). A 50% aqueous solution of ethylene oxide is the most irritating combination. The onset of the skin burn can be delayed. Skin sensitization in ethylene oxide plant workers who were challenged with a single dermal application of ethylene oxide was not noted (262). However, some evidence has been reported of contact dermatitis in a small proportion of exposed workers (263,264). Skin sensitization studies in guinea pigs are negative (265).

Swallowing ethylene oxide is a highly unlikely route of exposure. However, harmful effects, including coma, death, and severe irritation and ulceration of the mouth and throat, could occur.

OSHA considers that, at excessive levels, ethylene oxide may present reproductive, mutagenic, genotoxic, neurologic, and sensitization hazards. In addition, ethylene oxide is considered by OSHA, IARC, and NTP as a potential human carcinogen.

Explosibility and Fire Control. As in the case of many other reactive chemicals, the fire and explosion hazards of ethylene oxide are system-dependent. Each system should be evaluated for its particular hazards including start-up, shut-down, and failure modes. Storage of more than a threshold quantity of 5000 lb (~2300 kg) of the material makes ethylene oxide subject to the provisions of OSHA 29 CFR 1910 for "Highly Hazardous Chemicals." Table 15 summarizes relevant fire and explosion data for ethylene oxide, which are at standard temperature and pressure (STP) conditions except where otherwise noted.

Table 15. Fire and Explosion Hazard Evaluation Data

Property	Value
<i>Liquid properties</i>	
electrical conductivity (25°C), pS m	4 × 10 ⁶
reaction threshold temperature (ARC, 0.02°C min, ϕ 1), °C	~200
<i>Vapor properties</i>	
standard heat of formation, kJ mol ^a	52.63
heat of decomposition (max), kJ mol ^a	134.3
minimum decomposition pressure, kPa ^b	40
decomposition flame temperature (<i>P</i> constant), K	1277
decomposition pressure ratio (<i>V</i> constant)	~11.0
autodecomposition temperature, K	~773
fundamental burning velocity, m s	2.7 × 10 ⁻²
minimum ignition energy, J ^a	~1
heat capacity, <i>C_p</i> , J (mol·K) ^a	48.28
thermal conductivity, W (mol·K)	0.124
<i>Properties of vapor–air mixtures</i>	
lower flammable limit, mol %	3.0
upper flammable limit, mol %	100
deflagration <i>K_G</i> index ^c J, (MPa·m) s ^d	18.8 — 24.6
deflagration <i>P_{max}</i> index ^c , kPa	968 — 975
minimum deflagration pressure (~5–60 mol%), kPa ^b	1.3–2.6
stoichiometric composition, mol %	7.72
detonable range, mol %	5–30
recommended TNT efficiency, %	10
low heat of combustion, kJ mol ^a	1217
high heat of combustion, kJ mol ^a	1305
theoretical flame temperature (<i>P</i> constant), K	2402
theoretical deflagration pressure ratio (<i>V</i> constant)	9.91
autoignition temperature, K	718
fundamental burning velocity, m s	1.08
minimum ignition energy (10.4 mol %), J ^a	6 × 10 ⁻⁵

^a To convert J to cal, divide by 4.184.
^b To convert kPa to mm Hg, multiply by 7.5.
^c (20 — 1000 L spheres).
^d 1bar = 10² kPa = 10⁻¹ MPa.

Liquid Hazards. Pure liquid ethylene oxide will deflagrate given sufficient initiating energy either at or below the surface, and a propagating flame may be produced (266,267). This requires certain minimum temperatures and pressures sensitive to the mode of initiation and system geometry. Under fire exposure conditions, an ethylene oxide pipeline may undergo internal decomposition either by direct initiation of the liquid, or by formation and subsequent decomposition of a vapor pocket (190).

Liquid mists of ethylene oxide will decompose explosively in the same manner as the vapor. Burning rate increases with decreased droplet size. While the deflagration pressure ratio for ethylene oxide vapor is about 11 or less, liquid mist decomposition can give much greater pressures and very fast rates of pressure rise (190).

Liquid ethylene oxide under adiabatic conditions requires about 200°C before a self-heating rate of 0.02°C min is observed (190,191). However, in the presence of contaminants such as acids and bases, or reactants possessing a labile hydrogen atom, the self-heating temperature can be much lower (190). In large containers, runaway reaction can occur from ambient temperature, and destructive explosions may occur (268,269).

Ethylene oxide is an electrically conductive liquid that does not accumulate static electricity in grounded equipment. Static electricity can, however, accumulate in liquid mist produced by splashing and spraying. Although the vapor alone has a large minimum ignition energy, mixtures with oxidants such as air can be very sensitive to ignition (190).

Vapor Hazards. Ethylene oxide vapor can decompose explosively and propagate a decomposition flame when its pressure is greater than about 300 mm Hg (depending on temperature). To prevent decomposition, dilution using an inert gas (N₂ or CO₂) or an extinguishant such as a halocarbon have been used. The amount of dilution required depends on temperature, pressure, and the anticipated ignition source (270). It is found that considerably more inert component is required for long duration ignition such as by a flame than would be needed for a short duration spark or fused wire igniter (271).

Similarly, the minimum ignition energy of the vapor is found to be substantially less using inductance sparks of long duration than with capacitance sparks (271). This behavior is typical of many decomposition flames and some slow-burning gases in air. Ignition sources of high power tend to distort the flame kernel, and the energy cannot be absorbed as efficiently as from a longer-duration ignition source of optimized characteristics (272).

Safe dilution requirements can be given for the gas phase in a flammability diagram or equation (270,273). Alternatively, safe vapor dilution can be given in terms of the liquid storage conditions where allowance can be made for solubility of the inert gas in liquid ethylene oxide (273).

As an alternative to inert operation, either the enclosure can be designed to withstand the overpressure due to a decomposition, or the enclosure can be fitted with a deflagration suppression system (271). Inert operation is usually preferred. To prevent a decomposition flame propagating into an enclosure, special flame arrestors comprising tube bundles might be considered (190). It is essential that the design be tested under worst-case practical conditions, and anticipates "burn through" should a flame stabilize at the end of the arrestor. As the arrestor heats, its ability to prevent flame propagation may be compromised. This situation may be detected by temperature sensors, and appropriate action taken to remove the flame.

At a sufficient temperature, EO vapor will autodecompose. The minimum autodecomposition temperature decreases with increased volume and pressure (190,191). The vapor is susceptible to local catalysis by certain high surface area metal oxides such as γ-Fe₂O₃ and γ-FeO(OH). The catalyst may cause rapid local heating and vapor ignition under near-adiabatic conditions. The most rapid rate of self-heating occurs when conditions promote EO disproportionation rather than isomerization or polymerization. The catalysts may accumulate on polyethylene oxide or other polymers in a system, and may have migrated considerable distances from the original metal oxide source. These polymers may also trap significant quantities of EO, which serve as fuel for propagating decomposition in proximity to the catalyst.

Hazards of Mixtures with Air. Pools of liquid ethylene oxide will continue to burn until diluted with at least 22 parts of water by volume. This must be increased to about 100 parts water if the vapor is confined, such as in a sewer.

Mixtures of ethylene oxide with air are far easier to ignite and burn much faster than the pure vapor. Flames may propagate at low pressures, 1.3–2.6 kPa (10–20 mm Hg), over a wide range of compositions. The burning velocity (or K_G index) is much larger than that of propane, and under the provisions of NFPA 68 (274) there is no design basis for venting an optimum ethylene oxide deflagration in weak enclosures such as buildings. In the open air, partial confinement of vapor clouds may cause significant blast overpressures (190). A TNT equivalence of 10% has been suggested (275).

When ethylene oxide vapor or liquid leaks into porous refractory insulation such as mineral wool or calcium silicate, it reacts with water contained in the insulation forming low molecular weight poly(ethylene glycol)s. Whereas ethylene oxide is volatile, the glycols can accumulate and self-heat in the presence of air. Runaway self-heating requires temperatures above ~60° C and typically ~80° C (191). The event may lead to a fire in the insulation that might overheat small diameter lines causing internal decomposition. To prevent this, cellular glass insulation may be used since it is nonporous and cannot accumulate glycols (191).

Catalysts such as iron oxides cause isomerization of the ethylene oxide to acetaldehyde with the evolution of heat. The acetaldehyde has a much lower autoignition temperature in air than does ethylene oxide, and the two effects may lead to hot-spot ignition (190,191).

Uses

Ethylene oxide is an excellent fumigant and sterilizing agent (231). Ethylene oxide is used as an antimicrobial pesticide to fumigate spices and to sterilize medical devices, such as sutures, bandages, catheters, endoscopes, and cardiac pacemakers. Over half the medical devices made in the United States are sterilized using ethylene oxide; every hospital performing surgery has at least one ethylene oxide sterilizer. Ethylene oxide gas sterilants permit convenient sterilization of delicate instruments and supplies made of almost any material. Ethylene oxide readily penetrates deep pores and narrow crevices, and passes through wrappings of most polymers, paper, and cloth. The ethylene oxide sterilization process requires relatively low temperatures and pressures, and does not damage the materials or packaging being sterilized.

Ethylene oxide sterilant gases are supplied as liquified compressed gases, either pure or as a mixture with a flame retardant. When supplied as a pure gas, the ethylene oxide is shipped in special insulated containers. For safety reasons, nitrogen gas is added to the vapor phase up to a total pressure of 345 kPa (50 psig) at 21°C. When used in a sterilizing chamber, the flammability of ethylene oxide is usually controlled by purging the sterilization chamber with nitrogen gas at the beginning and the end of the sterilization process. In some cases, the effects of a potential deflagration are moderated by operating under great vacuum or, in the case of small hospital sterilizers, by using very small quantities of ethylene oxide. During 1960–1990, the most common flame retardants used for ethylene oxide mixtures were CO₂ and dichlorodifluoromethane (CFC-12). The CFC-12 mix (27.3 vol % ethylene oxide, 72.7 vol % CFC-12) accounted for over 90% of all sterilant gas used in the United States. With concern about the role of CFC-12 in causing depletion of stratospheric ozone, sterilizers and suppliers have committed to stop using this mixture. It will be replaced by more use of ethylene oxide nitrogen and by use of alternative flame retardants.

Ethylene oxide has been studied for use as a rocket fuel (276) and as a component in munitions (277). It has been reported to be used as a fuel in FAE (fuel air explosive) bombs (278).

Derivatives

Ethylene Glycol. Well over 50% of the ethylene oxide produced is used in the manufacture of ethylene glycol. Ethylene glycol [107-21-1] is used in two significant applications: as a raw material for poly(ethylene terephthalate) for use in polyester fiber, film, and containers, and as an automotive

antifreeze. Other important uses for ethylene glycol are in heat-transfer fluids, deicing fluids (aircraft, runway, and coal), latex paints (to provide freeze–thaw stability), natural gas dehydration, and many others (279) (see ANTIFREEZES AND DEICING FLUIDS; GLYCOLS).

Di-, Tri-, and Tetraethylene Glycols. These products are coproducts of ethylene glycol. Diethylene glycol is used as an intermediate for polyester resins, polyurethanes, and for tri- and tetraethylene glycols (by reaction with ethylene oxide). It is used for the drying of natural gas, as a component in antifreeze and deicers, and a variety of other applications. Triethylene glycol is primarily used as a liquid drying agent for natural gas. It is also used as a humectant and a solvent, and in the manufacture of plasticizers and unsaturated polyester resins. Tetraethylene glycol has an important value in the extraction of aromatic hydrocarbons from nonaromatic hydrocarbons (280) (see GLYCOLS).

Nonionic Surface-Active Agents. Approximately 14^o of the ethylene oxide consumed in the United States is used in the manufacture of nonionic surfactants. These are derived by addition of ethylene oxide to fatty alcohols, alkylphenols (qv), tall oil, alkyl mercaptans, and various polyols such as poly(propylene glycol), sorbitol, mannitol, and cellulose. They are used in household detergent formulations, industrial surfactant applications, in emulsion polymerization, textiles, paper manufacturing and recycling, and for many other applications (281).

Ethanolamines. These are produced by the reaction of ethylene oxide and ammonia (see ALKANOLAMINES). Approximately one-third of the production is used in detergents. Other applications include natural gas purification, cosmetics, metalworking, textiles, and chemical intermediates (282).

Glycol Ethers. These are made by reaction of ethylene oxide with various alcohols. They are used for solvents, detergents, brake fluids, and jet fuel deicing.

Poly(ethylene glycol)s. These relatively low molecular weight ethylene oxide polymers are made by reaction of ethylene oxide with water or glycols. Molecular weights range from 200 to 14,000 (42). Poly(ethylene glycol)s of molecular weight 200–600 are water-white liquids. The higher molecular weight polymers range in consistency from a soft white grease-like petrolatum to hard waxy solids. They are used in cosmetics, ointments, pharmaceuticals, and as a food additive. They are also used as water-soluble lubricants for molds, textile processing, and ceramics, and as a wood preservative (43,281).

Poly(ethylene oxide). High molecular weight ethylene oxide polymers are made by coordinative anionic polymerization (43,281) on alkyls or alkoxides of Group IIA or IIIA metals. Commercially available materials (Polyox, Union Carbide) are available in a molecular weight range of 90,000–3,800,000. These polymers have diverse applications. A very dilute (<100 ppm) solution in water greatly reduces viscous drag. Agriculturally, it is used to encapsulate seeds at regular spacings in a tape leading to regular plant spacing and elimination of thinning. It is also used for warp sizing, coagulation, and as a water-soluble packing material (43,281) (see POLYETHERS).

Other Derivatives. Ethylene carbonate, made from the reaction of ethylene oxide and carbon dioxide, is used as a solvent. Acrylonitrile (qv) can be made from ethylene oxide via ethylene cyanohydrin; however, this route has been entirely supplanted by more economic processes. Urethane intermediates can be produced using both ethylene oxide and propylene oxide in their structures (281) (see URETHANE POLYMERS).

BIBLIOGRAPHY

"Ethylene Oxide" in *ECT* 1st ed., Vol. 5, pp. 906–925, by R. S. Aries, Consulting Chemical Engineer, and H. Schneider, R. S. Aries & Associates; in *ECT* 2nd ed., Vol. 8, pp. 523–558, by H. C. Schultze, Union Carbide Corp.; in *ECT* 3rd ed., Vol. 9, pp. 432–471, by J. N. Cawse, J. P. Henry, M. W. Swartzlander, and P. H. Wadia, Union Carbide Corp.

1. A. Wurtz, *Ann.* **110**, 125 (1859); *Ann. Chem. Phys.* **55**, 433 (1859).
2. A. Wurtz, *Ann. Chim. Phys.* **69**, 335 (1863).
3. W. Bone and R. Wheeler, *J. Chem. Soc.* **85**, 1637 (1904).
4. L. Reyerson and L. J. Swearingen, *J. Am. Chem. Soc.* **50**, 2872 (1928).
5. R. Willstater and M. Bommer, *Ann.* **422**, 136 (1921).
6. **Ger. Pat. 168,291**; **Fr. Pat. 360,785** (1905–1906); **Brit. Pat. 21,941** (Oct. 4, 1906), J. Walter.
7. **Pat 729,952** (Mar. 27, 1931), and additions **41,254** (July 4, 1931), **41,724** (Sept. 10, 1931), **41,484** (Sept. 25, 1931), **41,810**, **41,811** (Apr., 1933); **Fr. Pat. 739,562** (Oct. 3, 1931), T. E. Lefort (to Societe Francaise de Catalyse Generalisee).
8. **Brit. Pats. 402,438** and **402,749** (Dec. 4, 1933), T. E. Lefort (to to Societe Francaise de Catalyse Generalisee); **U.S. Pat. 1,998,878** (Apr. 23, 1935), T. E. Lefort (to Carbide and Carbon Chemicals Corp.); reissued as **U.S. Pats. 20,370** (May 18, 1937), **22,241** (Dec. 29, 1942).
9. *Ethylene Oxide, Brochure F-ICD23*, Union Carbide Corp., Danbury, Conn., 1993.
10. C. J. Walters and J. M. Smith, *Chem. Eng. Prog.* **48**, 337 (1952).
11. D. W. Davidson and G. J. Wilson, *Can. J. Chem.* **41**, 1424–1434 (1963).
12. B. J. Zwolinski and R. Wilhoit in D. E. Gray, ed., *American Institute of Physics Handbook*, 3rd ed., McGraw-Hill Book Co., Inc., New York, 1972, pp. 316–342.
13. W. F. Giauque and J. Gordon, *J. Am. Chem. Soc.* **71**, 2176 (1949).
14. F. S. Bichowsky and F. D. Rossini, *The Thermochemistry of Chemical Substances*, Reinhold Publishing Corp., New York, 1936, p. 46.
15. F. H. Field and J. L. Franklin, *Electron Impact Phenomena*, Academic Press, Inc., New York, 1957.
16. R. Rein and co-workers, *J. Chem. Phys.* **45**, 4743 (1966).
17. G. B. Kistiakowsky and W. W. Rice, *J. Chem. Phys.* **8**, 618 (1940).
18. R. W. Gallant, *Hydrocarbon Process.* **46**(3), 143 (1967).
19. C. L. Yaws, *Physical Properties*, McGraw-Hill Book Co., Inc., New York, 1977, pp. 167–176.
20. O. Maass and E. H. Boomer, *J. Am. Chem. Soc.* **44**, 1709 (1922).
21. Ref. 19, pp. 8, 85, and 175.
22. Ref. 19, pp. 135 and 175.
23. Ref. 19, pp. 145 and 175.
24. D. R. Stull, E. F. Westrum, and G. C. Sinke, *The Chemical Thermodynamics of Organic Compounds*, John Wiley & Sons, Inc., New York, 1969, p. 419.
25. K. A. Kobe and R. E. Pennington, *Pet. Refiner* **29**(9), 135 (1950).
26. J. D. Olson, *J. Chem. Eng. Data* **22**, 326 (1977).
27. A. Seidell, *Solubilities of Organic Compounds*, D. van Nostrand Co., Inc., New York, 1941.
28. R. E. Parker and N. S. Issacs, *Chem. Rev.* **59**, 737 (1959).
29. R. Bonacorsi, E. Scrocco, and J. Tomasi, *J. Chem. Phys.* **52**, 5270 (1970).
30. G. Franking and co-workers, *Bull. Chem. Soc. Jpn.* **48**, 2769 (1975).
31. H. Fujimoto and co-workers, *Bull. Chem. Soc. Jpn.* **49**, 1508 (1976).
32. C. A. Coulson, *Valence*, Oxford University Press, New York, 1976, pp. 215–216.
33. L. H. Meyer, A. Saika, and H. S. Gutowsky, *J. Am. Chem. Soc.* **75**, 4567 (1953).
34. C. Hirose, *Bull. Chem. Soc. Jpn.* **47**, 1311 (1974).
35. A. Komornicki, F. Pauzat, and Y. Ellinger, *J. Phys. Chem.* **87**, 3847 (1983).
36. E. D. Simandiras and co-workers, *J. Am. Chem. Soc.* **110**, 1388 (1988).
37. D. N. Glew and N. S. Rath, *J. Chem. Phys.* **44**, 1710 (1965).
38. R. K. McMullan and G. A. Jeffrey, *J. Chem. Phys.* **42**, 2725 (1965).
39. R. A. Nelson and R. S. Jessup, *J. Res. Natl. Bur. Standards* **48**, 206 (1952).
40. A. Rosowsky in A. Weissberger, ed., *Heterocyclic Compounds*, Vol. 19, pt. 1, Wiley-Interscience, New York, 1964.
41. M. S. Malinovskii, *Epoxides and Their Derivatives*, Daniel Davey & Co., Inc., New York, 1965 (especially for references from Eastern European countries).

42. *Glycols, Brochure F-41515B*, Union Carbide, New York, 1978.
43. F. E. Bailey and J. V. Koleske, *Poly(ethylene Oxide)*, Academic Press, Inc., New York, 1976.
44. J. Farakawa and T. Saegusa, *Polymerization of Aldehydes and Oxides*, Wiley-Interscience, New York, 1963.
45. K. C. Frisch and S. L. Reegen, *Ring Opening Polymerization*, Marcel Dekker, Inc., New York, 1969.
46. T. H. Baize, *Ind. Eng. Chem.* **53**, 903 (1961).
47. J. Dale, G. Borgen, and K. Daasvatn, *Acta Chem. Scand.* **B28**(3), 378 (1974); **U.S. Pat. 3,928,386** (Dec. 23, 1975), J. Dale, G. Borgen, and K. Daasvatn.
48. J. Dale and K. Daasvatn, *J. Chem. Soc., Chem. Commun.*, 295 (1976); **U. S. Pat. 3,997,563** (Dec. 14, 1976), J. Dale and K. Daasvatn.
49. R. M. Izatt and co-workers, *Chem. Rev.* **85**, 271 (1985).
50. A. Lourenco, *Ann. Chem. Phys.* **67**, 275 (1863).
51. J. N. Bronsted, M. Kilpatrick, and M. Kilpatrick, *J. Am. Chem. Soc.* **51**, 428 (1929).
52. J. Koskihalio and E. Whalley, *Trans. Faraday Soc.* **55**, 815 (1959).
53. F. A. Long, J. G. Pritchard, and F. E. Stafford, *J. Am. Chem. Soc.* **79**, 2362 (1957).
54. A. M. Eastham and G. A. Latremouille, *Can. J. Chem.* **30**, 169 (1952).
55. P. O. I. Virtanen, *Suomen Kemistilehti* **34B**, 62 (1961).
56. H. J. Lichtenstein and G. H. Twigg, *Trans. Faraday Soc.* **44**, 905 (1948).
57. J. Koskihalio and E. Whalley, *Can. J. Chem.* **37**, 783 (1959).
58. L. Smith, G. Wode, and I. Widhe, *Z. Physik. Chem.* **130**, 154 (1927).
59. R. F. Goldstein, *The Petroleum Chemicals Industry*, John Wiley & Sons, Inc., New York, 1958, p. 352.
60. L. Knorr, *Chem. Ber.* **32**, 729 (1899).
61. Ref. 56, pp. 354–355.
62. Ref. 37, pp. 327–328.
63. C. C. Culvenor, W. Davies, and N. S. Heath, *J. Chem. Soc.*, 278 (1949).
64. H. Lemaire in M. J. Schick, ed., *Nonionic Surfactants*, Marcel Dekker, Inc., New York, 1966, pp. 177–179.
65. Ref. 37, p. 330.
66. Ref. 37, pp. 386–417.
67. N. G. Gaylord and E. J. Becker, *Chem. Rev.* **49**, 413 (1951).
68. M. S. Kharasch and O. Reinmuth, *Grignard Reactions of Nonmetallic Substances*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1954.
69. V. Grignard, *Bull. Soc. Chim. Fr.* **29**, 944 (1903); *Compt. Rend.* **136**, 1260 (1903).
70. Ref. 37, pp. 350, 432–440.
71. E. L. Gustus and P. G. Stevens, *J. Am. Chem. Soc.* **55**, 378 (1933).
72. W. Deckert, *Angew. Chem.* **45**, 559 (1932).
73. O. F. Lubatti, *J. Soc. Chem. Ind. (London)* **54**, 424T (1935).
74. W. Zeise, *Chem. Ber.* **66B**, 1965 (1933).
75. Ref. 56, p. 358.
76. M. S. Malinovski and N. D. Medjanzema, *J. Gen. Chem. USSR* **23**, 221 (1953).
77. Ref. 37, pp. 418–428.
78. Ref. 37, pp. 308–315.
79. Ref. 56, pp. 348–352.
80. C. R. Enyeart in Ref. 61, pp. 80–81.
81. Ref. 56, p. 357.
82. S. Jaffe in H. M. England and W. T. Beery, eds. *Proceedings, Second International Clean Air Congress*, Academic Press, Inc., New York, 1971, pp. 316–324.
83. H. Gbsten, L. Klasinc, and D. Maric, *J. Atmos. Chem.* **2**(1), 83–93 (1984).
84. S. A. Greene and L. J. Gordon, *Jet Propul.*, 798 (July, 1957).
85. F. A. Burden and J. H. Burgoyne, *Nature* **163**, 723 (1949); *Proc. Roy. Soc. London Ser. A* **199**, 328 (1949).
86. G. B. Shah, E. P. Chock, and R. G. Rinker, *Ind. Eng. Chem. Fundam.* **10**(1), 13 (1971).
87. Union Carbide Chemicals and Plastics Company, Inc., *Ethylene Oxide Producer's Technology Conference, Executive Summary*, Atlanta, Ga., Jan. 22, 1992.
88. F. G. A. Stone and H. J. Emeleus, *J. Chem. Soc.*, 2755 (1950).
89. R. Landau and R. E. Lidov in S. A. Miller, ed., *Ethylene: Its Industrial Derivatives*, Ernest Benn Ltd., London, 1969, p. 613.
90. Ref. 37, pp. 453–454.
91. Ref. 37, pp. 343–344.
92. J. Buddrus, *Chem. Ber.* **107**, 2050–2061 (1974).
93. Ref. 86, pp. 567–576.
94. Ref. 86, pp. 576–580.
95. J. F. Norris, *J. Ind. Eng. Chem.* **11**, 817 (1919).
96. I. Kiguchi, T. Kumazawa, and T. Nakai, *Hydrocarbon Process.* **55**(3), 69 (1976).
97. B. J. Ozero and J. V. Procelli, *Hydrocarbon Process.* **63**(3), 55 (1984).
98. G. O. Curme and F. Johnston, *Glycols, ACS Monograph No. 114*, Reinhold Publishing Corp., New York, 1952, Chaps. 2 and 5.
99. *Ethylene Oxide and Glycol Worldwide Supply/ Demand Report*, Petrochemical Consultants International, 1990.
100. *Monthly Petrochemical and Plastics Analysis*, Chemical Data Inc., Oct., 1991.
101. *Chem. Mark. Rep.*, (Feb. 17, 1992).
102. Ref. 88, pp. 521–563.
103. W. M. H. Sachtlar, C. Backy, and R. A. Van Santen, *Catalysts Review-Sci. Eng.* **23**(182), 127–149 (1981).
104. **Eur. Pat. Appl. 176,253 A1** (Apr. 2, 1986), (to ICI).
105. **Eur. Pat. Appl. 266,852 A1** (May 11, 1988), (to Shell Oil Co.).
106. **Eur. Pat. Appl. 266015 A1** (May 4, 1988) (to Shell Oil Co.).
107. P. B. Grant and R. M. Lambert, *J. Catal.* **92**, 364–375 (1985).
108. G. H. Twigg, *Trans. Faraday Soc.* **42**, 284, and 657 (1946); *Proc. Roy. Soc. (London) Ser. A* **188**, 92 (1946).
109. M. Gans and B. J. Ozero, *Hydrocarbon Process.* **55**(3), 73 (1976); B. DeMaglie, *Hydrocarbon Process.* **55**(3), 78 (1976); D. E. Field and co-workers, *Engineering and Cost Study of Air Pollution Control for the Petrochemical Industry*, Vol. 6, EPA Report No. EPA-450/3-73-006-f, Environmental Protection Agency, Washington, D.C., June, 1975.
110. **Ger. Pat. 2,300,512** (July 26, 1973), R. P. Nielsen and co-workers (to Shell International Research); **Ger. Pat. 2,448,449** (Apr. 30, 1975), P. A. Kilty (to Shell International Research).
111. **Belg. Pat. 638,319** (Oct. 8, 1964), R. S. Davis (to Halcon International).
112. **Dutch Pat. Appl. 6,502,927** (Sept. 23, 1964), (to The Dow Chemical Co.).
113. **Brit. Pat. 1,075,454** (July 12, 1967), (to Halcon International).

114. U.S. Pat. **4,012,425** (Mar. 15, 1977), R. P. Nielsen and J. H. LaRochelle (to Shell Oil Co.).
115. K. R. Barker and J. C. Zomerdiijk, *Eur. Chem. News Chemscope* **24**, 28 (Oct. 12, 1973).
116. R. Landau and co-workers, *Chem. Eng. Prog.* **64**(3), 17 (1968).
117. C. H. Chilton, *Chem. Eng.* **65**(15), 100 (1958).
118. *Hydrocarbon Process.* **54**(11), 145 (1975).
119. *Hydrocarbon Process.* **52**(11), 129 (1973).
120. P. W. Sherwood, *Chim. Ind. (Paris)* **70**, 1078 (1953); *Pet. Process.* **9**, 1592 (1954).
121. W. E. Vaughan and R. M. Goepp, Jr., *U.S. Dept. Comm. Off. Tech. Serv.*, PB Rep. 79607 (1947); FIAT Final Report 875, May 2, 1947.
122. U.S. Pat. **3,552,122** (Jan. 5, 1971), M. Parmegiani, S. D. Milanese, and O. Bellofatto (to Snam Progetti).
123. R. J. Ruff, *Chem. Eng. Prog.* **53**(8), 377 (1957).
124. U.S. Pat. **3,603,085** (Sept. 7, 1971), M. Parmegiani and S. D. Milanese (to Snam Progetti).
125. N. Y. Step, O. N. Dymont, and B. B. Chesnokov, *Khim. Prom. (Moscow)* **48**, 645 (1972).
126. U.S. Pat. **3,083,213** (Mar. 26, 1963), M. L. Courter (to Shell Oil Co.).
127. U.S. Pat. **3,878,126** (Apr. 15, 1975), E. G. Foster (to Shell Oil Co.).
128. U.S. Pat. **3,119,837** (Jan. 28, 1964), H. A. Kingsley and F. A. Cleland (to Shell Oil Co.).
129. E. T. McBee, H. B. Haas, and P. A. Wiseman, *Ind. Eng. Chem.* **37**, 432 (1945).
130. L. Y. Margolis and O. M. Todes, *Isv. Akad. Nauk SSSR Otd. Khim. Nauk*, 52, (1952); *Chem. Abstr.* **46**, 5413 (1952).
131. P. H. Emmett, *Catalysis*, Vol. 7, Reinhold Publishing Corp., New York, 1960, p. 246.
132. L. Y. Margolis, *Usp. Khim.* **28**, 615 (1959).
133. Brit. Pat. **1,329,252** (Sept. 5, 1973), D. Bryce-Smith and co-workers.
134. A. Ayame and co-workers, *Nippon Kagaku Kaishi* **8**, 1189 (1974).
135. O. Svajgl and co-workers, *Chem. Prum.* **22**, 493 (1972).
136. U.S. Pat. **2,605,239** (July 29, 1952), G. W. Sears (to E. I. du Pont de Nemours & Co., Inc.).
137. U.S. Pat. **2,831,870** (Apr. 22, 1958), W. J. McClements and B. E. Elliott (to Allied Chemical & Dye Corp.).
138. Brit. Pat. **754,593** (Aug. 8, 1956), (to N. V. de Bataafasche Petroleum Maatschappij).
139. U.S. Pat. **2,805,207** (Sept. 3, 1957), F. J. Metzger.
140. Brit. Pat. **811,828** (Apr. 15, 1959), (to The Dow Chemical Co.).
141. Brit. Pat. **501,278** (Feb. 21, 1939), (to N. V. de Bataafasche Petroleum Maatschappij).
142. U.S. Pat. **2,686,762** (Aug. 17, 1954), E. L. C. Tollefson (to National Research Council of Canada).
143. A. Cambron and W. A. Alexander, *Can. J. Chem.* **34**, 665 (1956).
144. Jpn. Pat. **46 40,256** (Nov. 17, 1971), S. Ishii and co-workers (to Nippon Shokubai Kagaku Kogyo).
145. U.S. Pat. **3,793,231** (Feb. 19, 1974), H. Bergmann and co-workers (to Huels).
146. Brit. Pat. **1,300,971** (Dec. 29, 1972), (to Huels).
147. U.S. Pat. **3,843,491** (Oct. 22, 1974), M. Piro and co-workers (to Snam Progetti).
148. P. Hariott, *J. Catal.* **21**, 56 (1971).
149. R. Wolf and co-workers, *Chem. Tech. (Berlin)* **14**, 600 (1962).
150. U.S. Pats. **2,279,469** and **2,279,470** (Apr. 14, 1942), G. H. Law and H. C. Chitwood (to Carbide and Carbon Chemicals Corp.).
151. J. Barthory and co-workers, *Proc. Conf. Appl. Phys. Chem* **2**, 279 (1971).
152. J. J. Carberry and co-workers, *J. Catal.* **26**, 247 (1972).
153. U.S. Pat. **3,962,136** (June 8, 1976), R. P. Nielsen (to Shell Oil Co.).
154. Ref. 104.
155. Ref. 105.
156. Ref. 106.
157. S. A. Tan, R. B. Grant, and R. M. Lambert, *J. Catal.* **106**, 54–64 (1987).
158. P. W. Sherwood, *Oil Gas J.* **39**, 80 (1957); **40**, 150 (1957).
159. L. Akella, Ph.D. dissertation, University of Florida, University Microfilms International, Ann Arbor, Mich., 1983.
160. Ref. 88, pp. 521–563.
161. G. K. Boreskov and co-workers, *Kinet. Katal.* **3**, 214 (1962).
162. Dutch Pat. Appl. **6,502,927** (Sept. 23, 1964), (to the Dow Chemical Co.).
163. U.S. Pat. **3,758,418** (Sept. 11, 1973), W. J. Leonard (to Shell Oil Co.).
164. Brit. Pat. **1,411,315** (Oct. 22, 1975), A. L. Gelbshtein (to the USSR).
165. Rom. Pat. **55,800** (Sept. 15, 1973), I. Opreacu and co-workers (to Institutul Petrochim).
166. Belg. Pat. **638,319** (Oct. 8, 1964), R. S. Davis (to Halcon International).
167. Brit. Pat. **748,957** (May 15, 1956), R. B. Egbert (to Chempatents Inc.).
168. U.S. Pat. **3,119,837** (Jan. 28, 1964), H. A. Kingsley and F. A. Cleland (to Shell Oil Co.).
169. Belg. Pat. **707,567** (June 5, 1968), D. Brown (to Halcon International).
170. U.S. Pat. **4,582,950** (Apr. 15, 1986), P. J. Busse and D. B. Taggart (to InterNorth).
171. F. L. W. McKim and A. Cambron, *Can. J. Res.* **27B**, 813 (1949).
172. L. M. Kaliberdo and co-workers, *Kinet. Katal.* **8**, 463 (1967).
173. Eur. Pat. Appl. **0200518-A2** (Nov. 5, 1986), B. J. Ozero (to The Halcon SD Group).
174. U.S. Pat. **4,879,396** (Nov. 7, 1989), B. J. Ozero (to Scientific Design Co., Inc.).
175. U.S. Pat. **4,769,047** (Sept. 6, 1988), R. F. Dye (to Shell Oil Co.).
176. U.S. Pat. **3,766,714** (Oct. 23, 1973), J. W. Cunningham (to Shell Oil Co.).
177. Eur. Pat. Appl. **181,273** (May 14, 1986), H. Neel and F. Delannoy (to Atochem SA).
178. U.S. Pat. **4,437,938** (July 17, 1981), V. S. Bhise and R. Hoch (to Halcon SD Co.).
179. U.S. Pat. **4,437,939** (Nov. 16, 1981), V. S. Bhise and R. Hoch (to Halcon SD Co.).
180. Jpn. Pat. **60 126,278** (July 5, 1985), to Mitsui Toatsu Chemical Co.).
181. USSR Pat. **1,570,743** (June 15, 1990), V. A. Kurbatov and co-workers (to USSR).
182. K. R. Barker and J. C. Zomerdiijk, *Eur. Chem. News Chemscope* **24**, 28 (Oct. 12, 1973).
183. Eur. Pat. Appl. **322,323** (June 28, 1989), F. Delannoy and G. Letray (to Atochem SA).
184. U.S. Pat. **3,3355,547** (Aug. 15, 1967), D. Garrett (to Halcon International).
185. Fr. Pat. **1,330,900** (May 20, 1963), Paire and co-workers (to Kuhlmann Co.).
186. U.S. Pat. **3,265,593** (Aug. 9, 1966), D. G. Leis (to Union Carbide Corp.).
187. U.S. Pat. **5,034,134** (Aug. 26, 1991), K. F. George and co-workers (to Union Carbide Corp.).
188. Jpn. Pat. **1,163,142** (June 27, 1989), H. Sugawara and T. Okawa (to Mitsui Toatsu Chemicals, Inc.).
189. U.S. Pat. **4,802,988** (Feb. 7, 1989), C. R. Bartels and J. Reale (to Texaco Inc.).
190. L. G. Britton, *Plant/Operations Progress* **9**(2) (Apr., 1990).
191. L. G. Britton, *Plant/Operations Progress* **10**(1) (Jan., 1991).

192. **Brit. Pat. 1,463,324** (Feb. 2, 1977), (to FarbWerke Hoechst AG).
193. **U.S. Pat. 4,822,926** (Apr. 18, 1989), R. F. Dye (to Shell Oil Co.).
194. **U.S. Pat. 3,904,656** (Sept. 9, 1975), S. E. Broz (to PPG Industries, Inc.).
195. **U.S. Pat. 3,083,213** (Mar. 26, 1963), M. L. Courter (to Shell Oil Co.).
196. B. DeMaglie, *Hydrocarbon Process.* **55**(3), 78 (1976).
197. R. Landau, *Pet. Refiner* **32**(9), 146 (1953).
198. **U.S. Pat. 3,993,673** (Nov. 23, 1976), C. H. McMullen (to Union Carbide Corp.).
199. A. G. Lee, *The Chemistry of Thallium*, Elsevier, New York, 1971.
200. **U.S. Pat. 4,021,453** (May 3, 1977), W. F. Brill (to Halcon International).
201. **Belg. Pat. 853,864** (Oct. 24, 1977), N. Rizkalla and A. N. Naglieri (to Halcon International).
202. **U.S. Pat. 4,058,542** (Nov. 15, 1977), N. Rizkalla and A. N. Naglieri (to Halcon International).
203. **Belg. Pat. 855,127** (Nov. 28, 1977), (to Halcon International).
204. **Ger. Pat. 2,541,526** (Apr. 1, 1976), A. P. Gelbein (to Lummus Co.).
205. **Eur. Pat. Appl. 84,301,864 9** (Nov. 7, 1984), R. L. Kelly (to Brit. Pet. Chemicals).
206. **U.S. Pat. 3,427,235** (Feb. 11, 1979), J. A. M. Le Duc (to Pullman Inc.).
207. **U.S. Pat. 3,288,692** (Nov. 29, 1966), J. A. M. Le Duc (to Pullman Inc.).
208. **U.S. Pat. 3,723,264** (Mar. 27, 1973), J. A. M. Le Duc and co-workers (to Pullman Inc. and Farbenfabriken Bayer Aktiengesellschaft).
209. A. Renken, M. Muller, and C. Wandrey, *Proc. 4th Intl. Conf. Chem. Reac. Eng.*, Heidelberg, 107, 1976.
210. **Ger. Pat. 2,605,991** (Aug. 25, 1977), A. Renken, C. Wandrey, and M. Muller (to Hoechst AG).
211. **U.S. Pat. 2,430,443** (Nov. 11, 1947), S. B. Becker (to Standard Oil Co.).
212. **U.S. Pat. 2,578,841** (Dec. 18, 1951), N. C. Robertson and R. T. Allen (to Celanese Corp.).
213. **U.S. Pat. 2,600,444** (June 17, 1952), F. W. Sullivan, Jr. (to GAF Corp.).
214. **U.S. Pat. 2,628,965** (Feb. 17, 1953), F. W. Sullivan, Jr. (to GAF Corp.).
215. **U.S. Pat. 2,684,967** (July 27, 1954), C. H. O. Berg (to Union Oil Co.).
216. T. E. Corrigan, *Pet. Refiner* **32**(2), 87 (1953).
217. **U.S. Pat. 3,948,609** (Apr. 6, 1976), N. M. Guseinov and co-workers (to the USSR).
218. **Brit. Pat. 1,379,797** (July 28, 1972), N. M. Guseinov and co-workers (to the USSR).
219. **USSR Pat. 562,555** (June 25, 1977), N. M. Guseinov and co-workers (to the USSR).
220. *Processing* **21**(11), 54 (1973).
221. F. B. Abeles and L. J. Dunn, *J. Plant Growth Regul.* **4**, 123–128 (1985).
222. **Eur. Pat. Appl. 83 303,695** (Jan. 11, 1982), C. H. Hou, A. I. Laskin, and R. N. Patel (to Exxon Research and Engineering Co.).
223. **Eur. Pat. Appl. 98,138** (Jan. 11, 1984), C. H. Hou, A. I. Laskin, and R. N. Patel (to Exxon Research and Engineering Co.).
224. J. A. M. DeBont and co-workers, *Enzyme Microb. Technol.* **5**(1), 55 (1983).
225. *Chem. Mark. Rep.*, Feb. 12, 1990.
226. International Agency for Research on Cancer, *Ethylene Oxide in LARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man* **11**, 157 (1976).
227. *U.S. Code of Federal Regulations*, Title 49, paragraph 173.323.
228. R. L. Anderson in F. D. Snell and L. S. Ettre, eds., *Encyclopedia of Industrial Chemical Analysis*, Vol. 12, John Wiley & Sons, Inc., New York, 1971, pp. 317–340.
229. E. J. Bond and T. Dumas, *J. Agric. Food Chem.* **30**, 986 (1982).
230. M. Collins and M. J. Barker, *Am. Lab.* **72** (1983).
231. *Gas Sterilants, Brochure L-2085-A*, Linde Division, Union Carbide Corp., Danbury, Conn., 1985.
232. M. Adams and M. Collins, *Anal. Proc.* **25**, 190 (1988).
233. L. Golberg, *Hazard Assessment of Ethylene Oxide*, CRC Press, Boca Raton, Fla., 1986.
234. K. H. Jacobson and co-workers, *AMA Arch. Ind. Health* **13**, 237–244 (1956).
235. Union Carbide Corp. internal laboratory report.
236. A. M. Thiess, *Archiv. Toxico.* **20**, 127–140 (1963).
237. W. M. Jay and co-workers, *Am. J. Ophthalmol.* **93**, 727–732 (1982).
238. P. F. Finelli and co-workers, *Arch. Neurol.* **40**, 419–421 (1983).
239. R. L. Hollingsworth and co-workers, *AMA Arch. of Ind. Health* **13**, 217–227 (1956).
240. D. W. Lynch and co-workers, *Toxicol. Appl. Pharmacol.* **76**, 69–84 (1984).
241. W. M. Generoso and co-workers, *Environ. Mutagen.* **8**, 1–7 (1986).
242. W. M. Generoso and co-workers, *Mutat. Res.* **199**, 175–181 (1988).
243. Osterman-Golkar and co-workers, *Mutat. Res.* **34**, 2 (1976).
244. L. Ehrenberg and co-workers, *Mutat. Res.* **24**, 83–103 (1974).
245. W. M. Snellings and co-workers, *Toxicol. Appl. Pharmacol.* **75**, 105–117 (1984).
246. *National Toxicology Program Technical Report*, National Institutes of Health publication number 88-2582, U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, Washington, D.C., 1987.
247. C. Hogstedt and co-workers, *J. Am. Med. Assoc.* **24**, 1132–1133 (1979).
248. C. Hogstedt and co-workers, *J. Am. Med. Assoc.* **255**, 1575–1578 (1986).
249. L. Hagmar and co-workers, *Ind. Arch. Environ. Health* **63**, 271–277 (1991).
250. M. J. Gardner and co-workers, *Br. J. Ind. Med.* **47**, 860–865 (1990).
251. H. L. Greenberg and co-workers, *Br. J. Ind. Med.* **47**, 221–230 (1990).
252. N. Kiesselback and co-workers, *Br. J. Ind. Med.* **47**, 182–188 (1990).
253. R. W. Morgan and co-workers, *J. Occup. Med.* **23**, 767–769 (1981).
254. P. L. Hackett and co-workers, *NTIS No. PB83-258038*, U.S. Department of Health and Human Services, Public Health Services, Centers for Disease Control, National Institute for Occupational Safety and Health, Washington, D.C., 1982.
255. W. M. Snellings and co-workers, *Toxicol. Appl. Pharmacol.* **64**, 476–481 (1982).
256. W. M. Snellings and co-workers, *Toxicol. Appl. Pharmacol.* **63**, 382 (1982).
257. J. W. Embree and co-workers, *Toxicol. Appl. Pharmacol.* **40**, 261–267 (1977).
258. R. B. Cumming and T. A. Michaud, *Environ. Mutagen.* **1**, 166–167 (1979).
259. K. Hemminki and co-workers, *Br. Med. J.* **285**, 1461–1463 (1982).
260. J. S. Taylor, *Dermatologic Hazards from Ethylene Oxide* **19**, 189–192 (1977).
261. R. J. Sexton and E. Henson, *J. Ind. Hyg. Toxicol.* **31**, 297–300 (1949).
262. A. M. Thiess, *Archiv. Toxico.* **20**, 127–140 (1963).
263. R. J. Sexton and E. Henson, *Ind. Hyg. Occup. Med.* **32**, 549–564, (1950).
264. J. L. Shupack and co-workers, *J. Lab. Clin. Med.* **98**, 723–729 (1981).

265. G. Woodard and M. Woodard, *Proc. Health Ind. Assoc. Tech. Symp.*, Washington, D.C., 1971, pp. 140–161.

266. B. Pesetsky and co-workers, *AIChE Loss Prevention* **13** (1980).

267. L. D. Chen and G. M. Faeth, *Combustion and Flame* **40**, 13–28 (1981).

268. J. E. Troyan and R. Y. LeVine, *AIChE Loss Prevention* **2**, 125–130 (1968).

269. R. G. Vanderwater, *Chem. Eng. Prog.* **85**(12) (Dec., 1989).

270. R. K. June and R. F. Dye, *Chem. Eng. Prog.* **85**(12) (Dec., 1989).

271. R. Siwek and E. Rosenberg, *Paper 52, 6th Int. Symp. Loss Prevention and Safety Promotion in the Process Industries*, Oslo, June 19–22, 1989.

272. L. G. Britton, *Plant/Operations Prog.* (Apr., 1992).

273. J. L. Brockwell, *Plant/Operations Prog.*, (Apr., 1992).

274. *Venting of Deflagrations*, NFPA 68, 1986 ed., National Fire Protection Association, Quincy, Mass.

275. *The Effects of Explosions in the Process Industries, Loss Prevention Bulletin 068*, Overpressure Working Party, Major Hazards Assessment Panel, Institution of Chemical Engineers, Apr., 1986.

276. B. Kit and D. S. Evered, *Rocket Propellant Handbook*, The MacMillan Co., New York, 1960, pp. 220–225.

277. **U.S. Pat. 3,940,443** (Feb. 24, 1976), C. A. Glass (to U.S.A. represented by the Secretary of the Navy).

278. *New York Post*, 5, (Apr. 25, 1975).

279. *Ethylene Glycol*, Brochure F-49193B-ICD, Union Carbide Corp., Danbury, Conn., 1991.

280. *Diethylene Glycol, Triethylene Glycol, Tetraethylene Glycol*; Brochures F-49181A-ICD, F-49191A-ICD, F-80066-ICD, Union Carbide Corp., Danbury, Conn., 1990.

281. F. E. Bailey and J. V. Koleske, *Alkylene Oxides and Their Polymers*, Marcel Dekker, Inc., New York, 1991.

282. *Ethanolamines*, Brochure F-41175C, Union Carbide Corp., Danbury, Conn., 1990.

J. P. Dever
K. F. George
W. C. Hoffman
H. Soo
Union Carbide Corporation

ETHYLIDENE DIACETATE.

See ACETIC ACID AND DERIVATIVES.

ETHYL MERCAPTAN (ETHANETHIOL).

See THIOLS.

ETHYNYLATION.

See ACETYLENE DERIVED CHEMICALS.

EUROPIUM.

See LANTHANIDES.

EUTROPHICATION.

See WATER, WATER POLLUTION.

EVAPORATION

Evaporation by its broadest definition is the conversion of a liquid to a vapor and applies to such widely diverse equipment as boilers, cooling towers, dryers, and humidifiers, and losses from fields, storage tanks, and reservoirs. In the narrower chemical engineering sense, evaporation is the removal of volatile solvent from a solution or a relatively dilute slurry by vaporizing the solvent. In nearly all industrial applications the solvent is water, and in most cases the nonvolatile residue is the valuable constituent. Evaporation differs from distillation (qv) in that when the volatile stream consists of more than one component no attempt is made to separate these components. Thus production of distilled water from impure feedwater utilizes an evaporator rather than a distillation unit even though small capacity units are usually called stills. Although the product of an evaporator system may be a solid, the heat required for vaporization of the solvent must be transferred to a solution or a slurry of the solid in its saturated solution in order that the device be classified as an evaporator rather than a dryer (see also DEWATERING; DRYING). It is not unusual for an evaporator to be used to produce a solid as its only product. For instance, table salt is produced by feeding a saturated brine to an evaporator, precipitating the salt as water is removed. A side stream of salt crystals in brine is withdrawn to a filter or centrifuge where the salt is recovered in essentially dry form; the filtrate is returned to the evaporator as a supplementary feed. Thus the heat required for evaporation of the water is transferred to a slurry in the evaporator even though the only material leaving the system is a solid, except for the evaporated water; usually a small bleed of brine is necessary to purge from the system the impurities entering with the feed brine (see CHEMICALS FROM BRINE).

The highly varied purposes for which evaporators are used industrially include: (1) reducing the volume to economize on packaging, shipping, and storage costs, eg, of salt, sugar (qv), caustic soda (see ALKALI AND CHLORINE PRODUCTS), orange juice (see FRUIT JUICES), and milk (see MILK AND MILK PRODUCTS); (2) obtaining a product in its most useful form, eg, salt from brine or sugar from cane juice; (3) eliminating minor impurities, eg, salt, sugar; (4) removing major contaminants from a product, eg, diaphragm cell caustic soda solutions contain more salt than caustic when produced but practically all the salt can be precipitated by concentrating to a 50% NaOH solution; (5) concentrating a process stream for recovery of resources, eg, pulp (qv) mill spent cooking liquor, if concentrated sufficiently in an evaporator, can be burned in a boiler to produce steam, yielding also an ash that can be used to reconstitute fresh cooking liquor; (6) concentrating wastes for easier disposal, such as nuclear reactor wastes (see NUCLEAR REACTORS), dyestuff plant effluents, and cooling tower blowdown streams; (7) transforming a waste into a valuable product, such as spent distillery slop after alcohol recovery, which can be concentrated to produce an animal feed (see FEEDS AND FEED ADDITIVES); and (8) recovering distilled water (qv) from impure streams such as sea water and brackish waters.

Salt was produced in prehistoric times from seawater or saline waters by solar evaporation in ponds. This method is still in widespread use and probably accounts for more total evaporation than any other process. The first artificially heated evaporators date to Roman times, again for salt, in flat pans over wood fires. Originally the pans were of lead, later of wrought iron, and such pans heated by coal were still in use in England in the 1960s (primarily to produce a salt of low bulk density for which there was a substantial export market). Similar open pans, heated by steam coils immersed in the brine, are used for production of salt having special grain characteristics in the United States. Evaporators were introduced by the sugar industry: the first steam-heated evaporator about 1800; the first one using vacuum in 1812; the first multiple-effect type in 1843; and the first vapor-compression evaporator about 1880. In the 1990s, the majority of evaporators are steam heated and use multiple-effect or vapor compression as the means of reducing the energy required for evaporation.

Steam-Heated Evaporators

The three principal requirements of steam-heated evaporators are: (1) transfer to the liquid of large amounts of heat needed to vaporize the solvent, (2) efficient separation of the evolved vapor from the residual liquid, and (3) accomplishing these aims with the least expenditure of energy justifiable by the capital cost involved (see ENERGY MANAGEMENT; HEAT EXCHANGE TECHNOLOGY). Most steam-heated evaporators use metal tubes as heat transfer surfaces, although some employ flat plates or, for special corrosion problems, impregnated carbon tubes. Final separation of the entrained liquid from evolved vapor

may be obtained in external centrifugal or mesh-type separators but preliminary separation is generally required (see SEPARATION, CENTRIFUGAL). An evaporator is frequently also used as a crystallizer, and therefore its design may be influenced by the need to produce crystals of the desired size, shape, and purity without adversely affecting the vapor–liquid separation and heat transfer functions (see CRYSTALLIZATION). Many types of evaporators are available to perform these functions. The choice depends on the characteristics of the liquor being handled with regard to heat transfer properties and tendencies toward salting, scaling, fouling, foaming, and corrosion. The extent to which it is desired to conserve heat also influences the choice of evaporator type.

Evaporator may refer either to the type of construction utilized or to the entire assemblage of equipment in a single installation. Thus a single multiple-effect evaporator may contain a number of effects of either the same or different evaporator types. An effect is a section of the evaporator heated by steam at one pressure and releasing vapor (water) at a lower pressure to another section. The term steam generally indicates the heat supply, whereas vapor means the material evaporated. Thus vapor from one effect becomes steam at the next effect. The term prime steam identifies the steam supplied from an outside source to operate the evaporator (see also STEAM). An effect may consist of several bodies, all operating at the same steam and vapor pressures. The purpose of more than one body in an effect may be to handle liquor at different concentrations, or the result of size limitations or of additions to increase the capacity of an existing evaporator.

Natural Circulation Evaporators. Natural circulation evaporators (Fig. 1) were the first developed commercially and still represent probably the largest number of units in operation. These evaporators utilize the density difference between the liquid and the generated vapor to circulate the liquid past the heating surface and thereby give good heat-transfer performance. The heat transfer tubes may be either vertical or horizontal, with the liquor either inside or outside the tubes. The horizontal tube type of Figure 1a has the heating steam inside the tubes that are immersed in the boiling liquid. This type was originally built with rectangular cast-iron bodies having a hemicylindrical top that required little floor space or headroom. The principal use is for making distilled water for boiler feed. These evaporators are incorporated in the power plant cycle, usually as single effects heated by turbine bleed steam and exhausting vapor to the feed heater circuit (see POWER GENERATION). They frequently operate under considerable pressure, and therefore employ horizontal cylindrical shells to better withstand the pressures and give a large liquid surface area for efficient disengagement. The type shown in Figure 1b was developed mainly for use in sugar mills, and was known as the standard evaporator; now it is known as the calandria or short-tube vertical (STV) evaporator. It employs fairly large (usually 0.05-m) diameter tubes only about 2 m long that are easily cleaned mechanically. Good heat transfer requires downtakes to permit recirculation through the tubes. The downtakes must have a flow area on the order of 60° or more of the flow area through the tubes themselves. The use of large-diameter short tubes and the need for a large downtake area result in a large body diameter relative to the amount of heating surface provided, usually larger than would be needed for vapor–liquid disengagement alone.

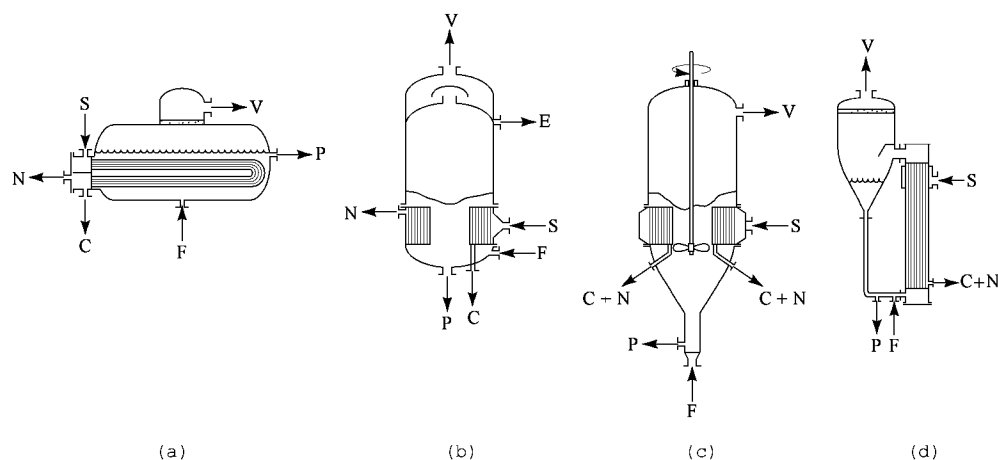


Fig. 1. Natural circulation evaporators where C = condensate, E = entrainment return, F = feed, N = noncondensables' vent, P = product or concentrate, S = steam, V = vapor, and  = knitmesh separator: (a) horizontal-tube, (b) short-tube vertical, (c) propeller calandria, and (d) long-tube recirculating.

The natural circulation types of Figure 1a and 1b are not generally suited for handling feeds that deposit appreciable amounts of solids. The reason is that circulation is by boiling action alone and any interruption in boiling, or even a reduction in boiling rate, allows the solids to drop out of suspension, increasing the tendency for subsequent deposition to occur on the heating surface rather than forming new solids in suspension. However, by adding a propeller in the downtake, as in Figure 1c, the solids can be kept better in suspension. This type, called the propeller calandria, has been in use since the early 1900s for crystallizing table salt from brine. Although it might be thought of as another version of the forced circulation evaporator, its heat transfer performance is about the same as the STV type of Figure 1b, indicating that the propeller is not really contributing directly to heat-transfer performance but only keeps salts from forming on the tubes, which would then impede heat transfer.

The natural circulation type shown in Figure 1d is called the long-tube recirculation (LTR) or recirculating LTV evaporator. It employs longer tubes in an external heating element that is more easily accessible for cleaning. In this case, the vapor-head size is determined only by disengagement requirements and can be appreciably smaller than those shown in Figures 1a–c. Smaller vapor-head size has a resultant shorter holdup time, and the easier cleaning makes these evaporators well-suited for degradable products such as milk.

Forced Circulation Evaporators. The forced circulation evaporator, suitable for the largest variety of applications, is usually the most expensive type. It usually consists of a shell-and-tube heat exchanger, a vapor–liquid separator (variously called vapor head, vapor body, separator, flash chamber, or body), and a pump to circulate the liquor from the body through the heater and back to the body. The system is usually arranged so that there is no boiling in the heater. The heat input is therefore absorbed as sensible heat, and vapor liberation does not occur until the liquor enters the flash chamber. Absorption of the heat input as sensible heat results in a temperature rise that reduces the net temperature difference available for heat transfer. To keep this temperature rise to reasonable limits, usually on the order of 2–6 K, requires circulating large volumes of liquor relative to the amount evaporated, eg, ca 0.11 m³ liquor per kilogram evaporated for dilute aqueous solutions at a 5 K rise. There is also an upper limit to temperature rise, usually on the order of 10 K, beyond which flashing at the entry to the flash chamber becomes so violent that large masses of liquor are ejected with the vapor. This makes entrainment separation more difficult and may impose structural shock loads on the separator. The head requirements of the circulating pump are generally quite low, consisting primarily of conventional friction and acceleration and deceleration losses at heater and body inlet and outlet, plus vortex losses in the body. The circulating pump is therefore usually of the propeller or mixed-flow type (see PUMPS).

The vapor body of a forced circulation evaporator is sized primarily for vapor–liquid separation and is arranged so that as much of the liquid as possible entering from the heater "sees" the pressure existing in the vapor space. When that does not occur, some liquid is recycled to the pump without flashing fully to the equilibrium pressure, resulting in another loss in ΔT available for heat transfer, which is variously called the submergence, short-circuiting, nonequilibrium, or Δ' loss. This loss can be minimized by returning liquor above the liquid level in the body, which adds to pump head, or by introducing the liquor tangentially, which sets up secondary circulation patterns in the body, to bring more of the heated liquid to the surface. If the evaporator is used as a crystallizer, the sizing of the vapor body may also be influenced by crystallization requirements. These may dictate the liquor volume in the system, crystal surface area available for growth, temperature rise through the heater, and physical arrangement of the body. If used as a crystallizer, discontinuities should be avoided in the walls near the operating level or in the splash zone, such as peepholes, manholes, and body flanges, on which salts can deposit. Furthermore, body walls should be so smooth that the salt does not adhere, or else so rough that it stays in place until cleaning time. The intermediate situation is to be avoided because this results in salt coming off the walls in thicknesses on the order of 0.01–0.1 m which can be transported by the circulating pump and deposited over the inlet ends of the tubes in the heater.

The heating element of a forced circulation evaporator is usually of the conventional shell-and-tube type, most often single pass and vertical but frequently multiple pass and horizontal. The heater is usually located far enough below the liquor level in the body so that hydrostatic head prevents boiling in the tubes. In crystallizing service it is desirable, but not always possible, to locate the heater far enough below the liquor level so that boiling does not occur even in a tube which has had its inlet blocked and thus has liquor in temperature equilibrium with the heating medium. This avoids complete filling of the tube with cemented solids which are most difficult to remove. Tube size and length are chosen to give reasonable (1.5–3- m s) tube velocities for the circulation rate available and the heating surface needed. In crystallizing service, small (less than 0.03-m) tube diameters are avoided to reduce risk of plugging, and the heaters are preferably vertical single-pass to afford more uniform distribution of flow to all tubes. Vortices in circulating lines and heater-inlet water boxes may result in such nonuniform velocities that there is little or no flow at all in some tubes and these quickly became plugged with solids.

Several configurations of forced circulation evaporators are shown in Figure 2. The most common arrangement is shown in Figure 2a having an external vertical single-pass heater and a tangential inlet to the body. Figure 2b shows the Oslo type of crystallizing evaporator in which the crystals are retained in a fluidized bed below the flash chamber. Because it is not always necessary to avoid boiling in the heating element, the element can project into the vapor head, as shown in Figure 2c. Auxiliaries shown in Figure 2 are not exclusive to each type of unit. Entrainment separators shown are top- and bottom-outlet centrifugal type and knit-mesh, respectively. The evaporator in Figure 2a includes a swirl breaker over the circulating pump inlet to reduce vortex losses in the vapor head and an elutriation leg to size, wash, and cool a crystallized product with part or all of the feed. In all cases, the steam inlet and vent outlet on the heat exchanger are placed so as to provide a positive flow path over the heating surface between the two. Forced circulation evaporators can be built for high single-unit capacities, having bodies as large as 15 m in diameter. The circulating pump is usually the limiting factor and it is not unusual to provide a single large body with as many as four or five separate heaters and circulating systems. For extreme fouling or salting conditions, the individual heaters can be arranged to be valved from the evaporator and cleaned without interrupting system operation.

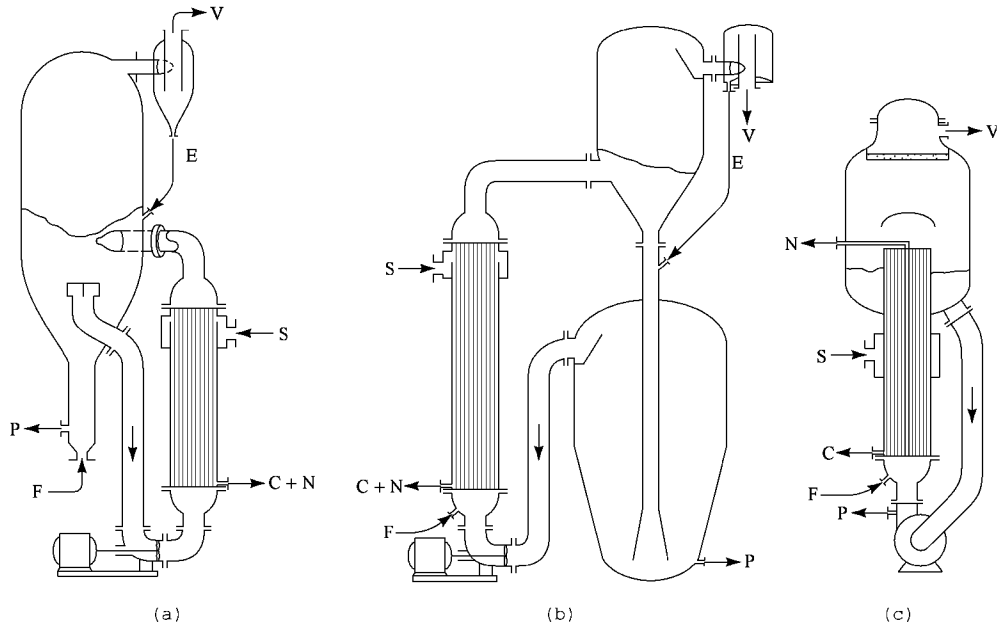


Fig. 2. Forced circulation evaporators: (a) submerged-tube, shown as circulating magma crystallizer; (b) submerged-tube, shown as suspension type crystallizer; and (c) boiling type. Terms are defined in Figure 1.

Film-Type Evaporators. Film-type evaporators are illustrated in Figure 3. Figure 3a shows the rising film or long-tube vertical (LTV) evaporator most widely used in the United States. It consists of a vertical shell-and-tube heat exchanger surmounted by a vapor–liquid separator. Another version uses an offset vapor head similar to Figure 1d. Tubes are generally 0.05 m or less in diameter and 6–10 m long, which permits packing a large amount of heating surface into a single shippable tube bundle (on the order of 3000 m²). Because of the simplicity of construction, costs per heating surface area are the lowest of any type, and heat-transfer performance is good under most operating conditions. Liquor flow through the LTV is generally once-through, distinguishing the LTV from the LTR of Figure 1d. However, if feed and discharge liquor properties differ widely, the evaporator may have partitions in the upper vapor head and lower inlet water box so that the effluent from one section can be returned to the inlet of another section of the total number of tubes in the bundle, and only the last group of tubes must handle liquor at the finished density. The LTV-type evaporator cannot handle crystallizing solutions but is excellent for foaming solutions because the deflector above the tube bundle acts as a foam breaker. Its widest use is for kraft-mill black liquor, which has foaming, fouling, and scaling tendencies and becomes quite viscous as it approaches discharge concentration (see PULP).

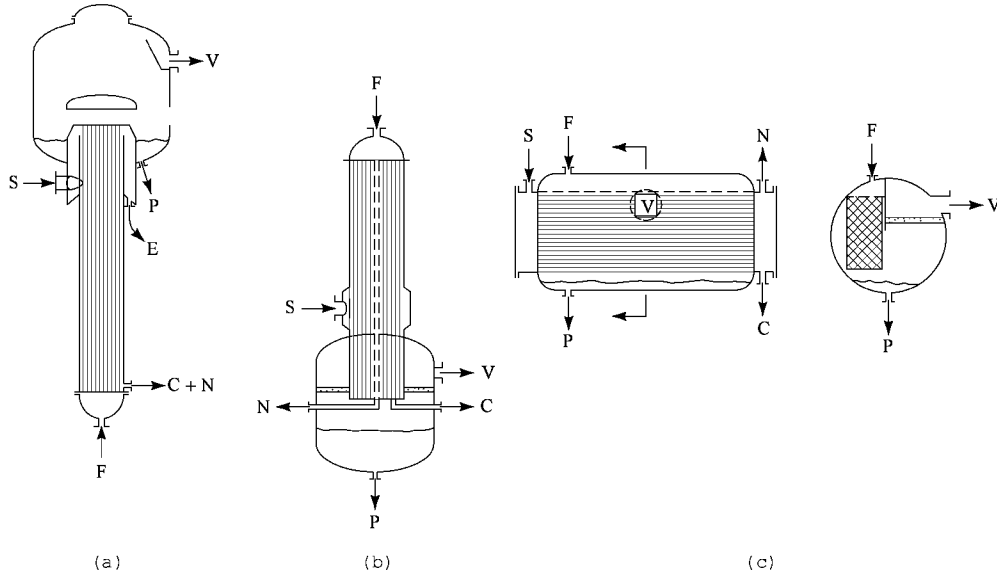


Fig. 3. Film-type evaporators: (a) long-tube vertical, (b) falling film, and (c) horizontal tube. Terms are defined in Figure 1. ☒ represents end view of (a).

The LTV is classified as a film-type evaporator because boiling takes place within the tubes and the vapor–liquid mixture is usually in the annular or film-flow regime for much of the tube length. High vapor velocities are generated and the interfacial shear also causes the liquid to move at high velocity, yielding good heat-transfer coefficients. However, frictional pressure drop, acceleration head developed as the vapor is generated, and hydrostatic head of the vapor–liquid mixture all cause the pressure and hence boiling temperature to be higher within the tube than at the tube exit, the temperature and pressure increasing with distance down from the outlet. Thus even if the feed is at its boiling point at the pressure in the vapor head, there is a section in the lower part of the tubes where the liquid cannot boil and is traveling at low velocity, and thus has poor heat-transfer characteristics. Proper selection of tube dimensions helps to minimize the adverse effects of reduced available ΔT and poor heat transfer in the non-boiling zone, as does the use of preheaters on the feed. These preheaters may be either external or incorporated as a part of the main tube bundle. Because LTV performance depends on the vapor–liquid velocities developed in the boiling zone, it follows that heat-transfer coefficients are a function of the load or of the overall ΔT . Coefficients are low at low temperature differences and low temperatures. This type of evaporator sometimes suffers from instability problems when operated under part-load conditions.

The falling-film evaporator shown in Figure 3b is an inverted version of the LTV that greatly reduces the adverse effects of pressure drop on available ΔT exhibited by the rising-film LTV. The hydrostatic head loss is eliminated, acceleration losses are lower because the liquid film is not accelerated substantially by the vapor flow, and the frictional pressure drop is generally only a little more than that of vapor flowing alone in a dry tube. In addition, heat-transfer performance is practically the same regardless of whether or not the film is boiling. This eliminates a poorly performing nonboiling zone at the inlet end of the tubes, even when the feed enters far below its boiling point. Uniform feed distribution to all tubes is usually the principal difficulty. Methods to overcome this include individual distributors in each tube, a full cone spray nozzle in the upper water box, a perforated plate above the top tube sheet with holes allowing feed to impinge on the tube sheet web between tubes, or, for high flow rates, no distribution devices at all. Feed rates are generally on the order of 1.5–8 m³ (h·m) of wetted tube perimeter. Because these are usually higher than the net feed rate, some recirculation is required. Thus the falling-film evaporator also belongs to the forced circulation class. As with the LTV, the inlet water box and the vapor head may be partitioned, permitting evaporation of the feed in stages; only the last stage has to operate at final density. This is advantageous when viscosity or boiling point increases rapidly as discharge concentration is approached but requires separate pumps for each stage.

The principal advantages of the falling-film evaporator are good heat-transfer performance, even at low temperature and low temperature differences, low initial cost, and excellent vapor–liquid separation characteristics. Principal applications have been for citrus juices, where performance at low temperature and low holdup is important, and applications requiring operation at low temperature differences, such as vapor compression or multiple-effect evaporators needing a large number of effects to be economical, eg, for producing fresh water from saline waters. The falling-film evaporator is normally not suitable for the usual crystallizing operations but has been very effective in scaling environments, scaling being avoided by maintaining a suspension of the scaling ingredient in the circulating liquid (1).

A special heating surface developed for this type of evaporator is the doubly fluted tube (2), which gives two to three times the heat-transfer coefficient of a smooth wall tube. On the steam side, the condensing coefficient is enhanced by the presence of longitudinal grooves. Surface tension forces draw the condensate into the grooves, leaving the area between the grooves bare so that they give coefficients comparable to those achieved with dropwise condensation. Heat-transfer coefficients on the boiling liquor side are also markedly higher than they are on smooth tubes. The principal application for these tubes is for production of distilled water from seawater (see WATER).

A combination of rising-film–falling-film evaporator utilizes a two-pass vertical heating element situated above the vapor head. The first pass, fed at the bottom, operates as a rising film and the vapor–liquid mixture then goes through the second pass as a falling film. This type evaporator usually operates with recirculation, but the amount of recirculation is less than in a normal forced circulation or falling-film evaporator. It was developed primarily for handling high viscosity fluids and for applications requiring low residence times.

Another version of the film evaporator is that in which liquid is showered over the outside of substantially horizontal tubes as shown in Figure 3c. This was originally called the Lillie evaporator and is now termed the horizontal-tube or spray-film evaporator. It has essentially the same advantageous heat-transfer characteristics as the falling-film type of Figure 3b and in addition requires much less headroom. Its main disadvantages are poorer vapor–liquid separation and greater difficulty of cleaning fouled tubes. Uniform liquor distribution at the top of the tube bundle is usually accomplished by perforated troughs or spray nozzles. Uniform distribution may not prevail further down in the bundle as vapor release from the side or ends of the bundle tends to drive liquid with it instead of the liquid falling vertically from one tube row to the next. In the original Lillie and in some modern versions, the tubes are rolled into a tube sheet at only one end and are sloped uphill so that condensate can drain countercurrent to the steam. The other end is then fitted with a perforated plug to act as a vent. Other versions employ tube sheets at both ends and frequently have several steam side passes, with a smaller number of tubes in each succeeding pass, to provide a tapered flow path from steam inlet to vent outlet.

The wiped-film or agitated-film evaporator, shown in Figure 4, uses mechanical energy to promote heat transfer. It employs a single large-diameter straight or tapered tube as the heating surface, in which a set of blades is rotated. The axis of rotation is frequently vertical. The blades maintain either a fixed close clearance from the wall or actually ride on the film of liquid and help to carry the liquid as a film around and along the length of the heating surface. The cost of these evaporators per unit of heating surface is very high and the capacity is relatively low, because a maximum of only about 40 m² of surface can be provided in a single tube. Thus, these evaporators are used primarily only for materials that cannot be handled in other evaporators, such as highly viscous liquids or liquids requiring very low residence times in contact with the heating surface. Because of the high capital cost and limited heating surface, these units are usually operated as a single effect at high ΔT s. This and structural considerations require that the large diameter tubes be of fairly heavy wall construction. Such evaporators exhibit poor heat-transfer performance on low viscosity fluids because of the added resistance of the metal wall.

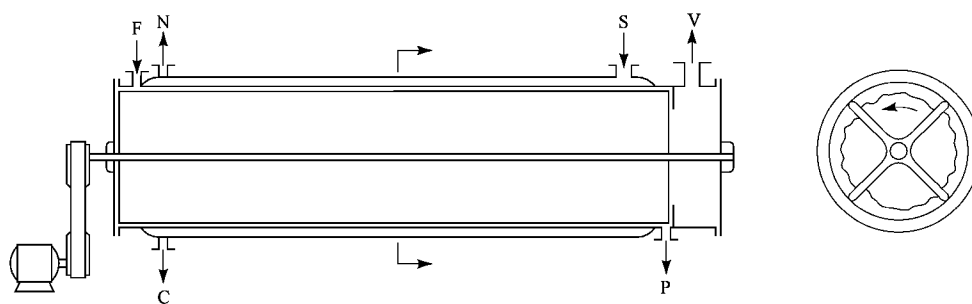


Fig. 4. Wiped-film evaporator. Terms are defined in Figure 1.

Energy Conservation

Most of the complexity and cost of an evaporator installation is a result of attempts to reduce energy consumption, which is usually by far the most important element of operating cost. Evaporators are not normally rated directly in efficiency of energy usage because the separation of solvent from the solution requires very little theoretical energy in an ideal system. Thus, for separating water from a salt solution having a boiling point rise of 5 K at its atmospheric boiling point, the minimum theoretical work energy requirement is only 30.1 kJ/kg (13.0 Btu/lb) of water removed, whereas it takes 2250

kJ/kg (970 Btu/lb) of heat to vaporize the water. Heat and work energy are not directly interchangeable in utility or in cost. Thus steam-heated evaporators are generally rated in terms of steam economy, ie, weight of water evaporated per weight of steam used, also called gained output ratio or performance ratio, and frequently standardized at pounds of water evaporated per 1000 Btu extracted from the steam (kg evaporated per 2324 kJ). Even evaporators that do use work energy (electrical or mechanical) for operation are not usually rated in terms of efficiency but instead in such terms as J/m³ (1.05×10^{-6} kWh/1000 gal) evaporated.

The single-effect evaporator is the simplest arrangement. It uses steam from an outside source and exhausts its vapor to the atmosphere or to an air- or water-cooled condenser. Such an evaporator requires about a 1-to-1 ratio of steam to water evaporated, but somewhat more if the feed is colder than the product and heat cannot be recovered by preheating the feed with concentrate and condensate. The high steam consumption limits the use of single effects to small capacities and to materials requiring an expensive type of evaporator, such as the wiped-film type, or having a very high boiling point, or to cases where the vapor is contaminated by materials that would cause excessive fouling or corrosion of heating surfaces when condensed, eg, rayon spin-bath liquor. Such evaporators may be operated on a continuous, batch, or semibatch basis with very little difference in heat requirements. In both batch and semibatch operation, final concentration is not reached until the end of the cycle, and these methods are therefore used primarily when the heat-transfer properties become markedly poorer as final concentration is approached. Semibatch operation is the more common. Feed is added continuously during most of the cycle in order to maintain a liquid inventory large enough to permit the evaporator to operate properly.

The single-effect evaporator produces almost as much vapor as the amount of steam used, the only difference being that the vapor is at a lower temperature and hence lower pressure. Compressing the vapor for reuse as the heat source was put into operation in the nineteenth century. This thermocompression or vapor recompression operation can be accomplished by either mechanical or steam-jet compressors. Mechanical compressors are by far the more efficient and may be driven by electric motors, gas or diesel engines, or steam or gas turbines. Because of the high specific volume of water vapor, positive displacement compressors are suitable for only the smallest capacities. Centrifugal compressors are most frequently used, whereas axial-flow multistage compressors are needed for the highest capacities, ca 50–500 m³/s (10³–10⁴ ft³/min). Compressor efficiencies are usually in the range of 70–75% for single-stage centrifugal machines (for compression ratios to about 1.75) and 80–85% for axial-flow machines (for compression) ratios of about 1.15 per stage). Steam-jet compressors, on the other hand, are only about 25–30% efficient. Therefore, if high pressure steam is available and capacity requirements are appreciable, a mechanical compressor driven by a steam turbine is preferred to a steam jet.

The ideal mechanical power requirement of a thermocompression evaporator is given by the Carnot equation:

$$W = Q \Delta T / T$$

where W is the work done, Q is the heat received at absolute suction temperature T , and ΔT is the difference in saturation temperature at compressor discharge and suction pressures. To minimize the work required ΔT is kept low so that the evaporator must have a large heating surface area. The optimum balance between power consumption and evaporator cost is usually at net ΔT in the range of 3–10 K. The need to operate at low net ΔT is a disadvantage of some types of evaporator, such as natural circulation and LTV units, where coefficients fall off at low ΔT ; and the submerged, forced circulation types, which lose ΔT as a result of temperature rise in the heating element and short-circuiting in the vapor head. The most advantageous evaporator type for vapor compression operation, when suitable for the liquor handled, is the falling film, which has very little ΔT loss (3). Because the ΔT across which the compressor must work also includes the boiling point rise (BPR) and any losses owing to pressure drop in vapor circuits, special care is exercised to minimize these. If the BPR of the product is high, multiple stages on the liquor side are advantageous so that only the last portion of the evaporation occurs at final product concentration. Vapor-side ΔT losses are minimized primarily by adequate duct sizing and the use of an entrainment separator having a low pressure drop, such as knitted mesh. Entrainment separation is particularly critical because the vapor becomes superheated on passing through the compressor and any liquid carried with it evaporates, depositing its contained solids on the blades. Entrainment should preferably be held to 5 ppm total solids or less to minimize compressor problems.

A thermocompression evaporator is like a flywheel because the compressor keeps a large amount of heat content circulating. However, heat input must balance heat output in order that the total energy content or temperature of the system remain constant. The work input of the compressor ultimately appears as a heat input to the system, and ideally this should balance the heat losses from the system so that no supplementary source of heat is needed, except at startup. The heat input from the compressor is generally quite small; thus it is necessary to preheat the feed very close to the evaporator temperature by heat exchange with condensate and concentrate. The approach temperature differences required in these feed heaters, in order to achieve a balanced system, decrease with decreasing evaporator ΔT and input power. Making a thermocompression evaporator more efficient requires increases in both evaporator and preheater heating surface area. Where possible, supplementary sources of makeup heat are utilized, such as waste-heat boilers on gas turbine drives or diesel engines. If steam-turbine driven or steam-jet compressors are used, a surplus of heat is usually available and it is necessary to bleed vapor at the compressor suction to maintain the evaporator at constant temperature.

The largest number of vapor-compression evaporators have been built for producing potable water from seawater or brackish water. These are usually relatively small capacity units for military use, offshore drill rigs, marine vessels, and the like, and use engine-driven compressors and short-tube vertical evaporators. Makeup heat is obtained from engine exhaust and engine cooling water. The principal advantage is the ability to build a complete, compact, highly efficient, easily transportable unit that needs only connection to fuel supply and a feedwater source for its operation. Another use has been for citrus juice concentration, which must be done at low temperatures to avoid degradation (see FRUIT JUICES). The need for low temperatures prevents use of multiple-effect evaporators for heat economy but also causes problems for vapor-compression operation because of the very high specific volume of steam under these conditions. A refrigerant, such as ammonia or Freon, can be employed as a secondary working fluid. The evaporator is then heated by the compressed refrigerant and the evaporator vapor is condensed in the refrigerant reboiler (see REFRIGERATION AND REFRIGERANTS). Thus this system has the disadvantage that the heat must be passed through two heating surfaces instead of one. Since the 1973 energy crisis, and as a result of higher efficiency, thermocompression evaporators are used in fields where they were not previously considered, such as for paper-mill waste cooking liquors (4) and for disposal of cooling-tower blowdown wastes. In the latter application, the cooling-tower water is usually nearly saturated with calcium sulfate, CaSO₄, and thus the evaporator must handle a severely scaling liquor. The falling-film evaporator has been used successfully in this case; seeding with CaSO₄ solids can prevent scale formation completely (1). Other uses include applications where lack of need for a heat sink is important and as a preconcentrator to increase the capacity and efficiency of existing multiple-effect installations. Diaphragm-cell caustic soda, for instance, is normally concentrated in forced circulation evaporators because of the large amounts of salt precipitated. However, an appreciable amount of water can be removed from the cell liquor before saturation is first reached with salt and the BPR is still relatively low; therefore, a falling-film thermocompression preconcentrator is advantageous.

If steam instead of power is the source of energy, multiple-effect evaporation is the principal means of energy conservation. In this operation, the vapor from one effect is used to heat another effect boiling at a lower temperature and the vapor from this effect is used to heat yet another effect boiling at still lower temperature. In such evaporators it is desirable to use an initial steam temperature as high as possible in the first effect and a heat sink temperature as low as possible to condense the vapor from the last effect, in order to develop the highest practical total temperature difference for heat transfer. The upper temperature limit is usually set by available steam pressure or scaling, fouling, product degradation, or corrosion characteristics of the liquor. The lower temperature limit is determined either by temperature and availability of cooling water or the need for low temperature steam for some other use. Other factors that may affect the lower temperature limit, which usually corresponds to a high vacuum, are poor heat transfer for most evaporator types at low temperature, high liquor viscosity, the very high vapor volumes from which liquor must be separated, and the cost of removing noncondensibles from the system.

Because each effect of an evaporator produces almost as much vapor as the amount it condenses, the total evaporation accomplished per unit of prime steam, or steam economy, increases in almost direct proportion to the number of effects used. The total heat load is also split up between the effects so that each effect has a much lower heat duty than a single effect for the same total evaporation load. However, the total available ΔT is also split up similarly so that each effect of a multiple effect requires about as much heating surface as a single effect operating over the same total temperature difference. Thus in selecting the number of effects to use in any installation, steam cost savings and capital cost of effects have to be balanced. Even before

the energy crisis of the 1970s, evaporators having six or seven effects were common in the pulp and paper (qv) industry. As many as 17 effects had been used in large seawater evaporators.

The relationships between number of effects, steam economy, and heating surface needs are not exact and can only be determined for a particular project by detailed heat and material balances, as well as consideration of the effects of temperature level and temperature difference on heat-transfer performance. The steam economy is generally not equal to the number of effects, primarily because of the influence of sensible heating loads. It frequently can be improved by alterations in feed sequence or installation of intereffect heat exchangers. The steam economy is usually close to a fixed fraction of the number of effects. The fraction may be over 1.0 when the feed is hot and may be as high as 0.9 even when the feed is cold, provided efficient means are included for preheating feed and for recovering waste heat at the least practical temperature differences. The most common method of heat recovery is by flashing the condensate produced in each effect to each of the lower temperature effects in turn so that the flash vapor is added to evaporator vapor to accomplish more useful evaporation. Prime steam condensate is usually not flashed because the flash vapor would be working at lower steam economy (over fewer effects) than the prime steam, and the condensate would then have to be reheated in the boiler circuit at the same efficiency as generating latent heat in the prime steam.

Similarly, heating surface area needs are not directly proportional to the number of effects used. For some types of evaporator, heat-transfer coefficients decline with temperature difference; as effects are added the surface needed in each effect increases. On the other hand, heat-transfer coefficients increase with temperature level. In a single effect, all evaporation takes place at a temperature near that of the heat sink, whereas in a double effect half the evaporation takes place at this temperature and the other half at a higher temperature, thereby improving the mean evaporating temperature. Other factors to be considered are the BPR, which is additive in a multiple-effect evaporator and therefore reduces the net ΔT available for heat transfer as the number of effects is increased, and the reduced demand for steam and cooling water and hence the capital costs of these auxiliaries as the number of effects is increased.

In a multiple-effect evaporator, the effects are numbered in the direction of steam flow, the first effect being the one heated by prime steam. Liquor feed sequence through the evaporator may be forward, backward, parallel, or mixed. Backward feed (to the coolest effect first and then successively through the higher temperature effects) is generally used when the feed is cold, because only a small volume has to be heated to the highest temperature, thereby reducing sensible heating losses and improving steam economy. Forward feed is generally used when feed is hot and when the concentrated product is not too viscous at the last effect temperature. Where necessary, forward feed can be used on a cold feed and can give almost the same steam economy as backward feed if the feed is preheated in stages by vapor extracted from each higher temperature effect of the evaporator in turn. Such an arrangement is used for seawater, which can be concentrated only in small amounts at high temperature without scaling but about threefold at the lowest effect temperature. These preheaters add to the complexity of the evaporator but not necessarily to total heating surface needs because they reduce the heat loads in the effects themselves. Parallel feed is frequently used in crystallizing operations and involves feeding to and withdrawing product from each effect. Mixed feed operation is common if feed is at some temperature intermediate between first and last effect, the finished product is too viscous to handle at low temperature, or liquor at an intermediate concentration and temperature is desired for further processing. All such conditions prevail in a kraft-mill liquor evaporator (see PULP). The evaporators are usually of the LTV type and the feed heaters are an integral part of the evaporator tube bundles. The highest temperature effect is frequently subdivided so that only a part of the tubes must work on liquor of the highest concentration and viscosity. Other flow-sheet variations include evaporators with several bodies in parallel on steam and vapor but in series on feed. Thus the kraft-mill evaporator might be an eight-body seven effect with two bodies in parallel in the last effect position in order to better handle the very large vapor volumes generated at the lowest temperature. Variations may also involve the steam path; eg, a combination double-triple effect, which may have one effect on prime steam with the evolved vapor being split, one part to a single effect and the other to a double effect. This variant is used when the liquor has a high BPR as it approaches final concentration, so high that not all evaporation can be accomplished, in this example, in a triple effect. Other flow sheets have been developed for specific types of evaporator, such as parallel split feed, which is used for desalination evaporators arranged in vertical stacks. In this case, about half the feed goes to the odd-numbered effects in one stack and the other half to the even-numbered effects in a second stack next to the first, with vapor connections crossing from one stack to the other.

Another means of gaining multiple-effect steam economy is by multistage flash operation, developed originally as the Alberger salt process. In this process, cool liquor is preheated in stages by vapor condensing at successively higher temperatures in a recovery section and finally by prime steam in a brine heater. The heated liquor is then flashed down in stages to generate the vapor used for preheating. The condensate from this vapor is also flashed down in the same manner so that the total flow being flashed is the same as the flow being heated. As a result, the temperature rise and flashing range in all heaters and flashers is about the same. Flashing through a 75 K range evaporates only about 12% of the feed; thus it is necessary to recirculate the liquor through the heaters and flashers when an appreciable degree of concentration is required. This then requires a rejection section in the flash train, containing one or more stages in which liquor is flashed to reject heat to cooling water before being recirculated to the recovery section. Even when recovering fresh water from seawater, recirculation is usually employed to reduce the amount of fresh feed that must be treated for scale prevention and deaerated before introduction into the evaporator. The characteristics of a flash evaporator are entirely different from those of a multiple-effect evaporator. Many more stages than effects are needed to achieve the same steam economy. Whereas in a multiple-effect evaporator the amount of heating surface provided is the most important determinant of capacity, and only the number of effects has a strong influence on steam economy, in a multistage flash evaporator the production rate is proportional only to the product of liquor circulation rate and total flash range. The steam economy is affected by both the number of stages and the amount of heating surface provided. The principal advantages of the multistage flash evaporator are that only one circulating pump is needed and that a number of stages, comprising both flash space and heating elements, can be combined in one vessel. The evaporator is in essence a number of forced circulation evaporator heating elements connected in series on one circulating pump, plus a number of forced circulation flash chambers connected in series. Because all heat is absorbed as sensible heat, the multistage flash evaporator suffers from the same loss in ΔT as the forced circulation evaporator, and also has short-circuiting losses of the same magnitude. The disadvantages of the multistage flash evaporator are therefore primarily this loss in net ΔT available for heat transfer, plus the fact that all liquor in the system is close to discharge concentration.

Although the interrelationships of the variables in a multistage flash evaporator is complex, the sizing of the heating surfaces is relatively simple, using the following factors (5): (1) heat-transfer coefficients from conventional correlations for condensing vapor and for liquid heating in tubes; (2) heat load, excluding brine heater: $Q = D\Delta H (1 - \Delta t / 1250)$; (3) weight of brine circulated: $W = Q / C_p \Delta t$; and (4) net ΔT for heat transfer $\Delta t = N \ln (1 - \Delta t / 1250 - R \cdot P \Delta t) / (1 - \Delta t / 1250 - R \cdot P \Delta t - P N)$, where Q = amount of heat transferred through heating surfaces; D = distillate production rate; ΔH = average latent heat of evaporation; Δt = total flashing range, K; C_p = specific heat of circulating brine; N = number of flash stages; R = average sum of BPR and short-circuiting losses; and P = steam economy desired based on 2324 kJ/kg (1000 Btu/lb) latent heat. From these, it is observed that heating surface can be traded against number of flash stages but that the number of stages should be on the order of three times the steam economy in order not to seriously decrease the net Δt available for heat transfer. Furthermore, the ratio of heating surface needed to the amount of brine circulated is usually so high that the recirculating liquid must be pumped through long lengths of small-diameter tubing in order to achieve reasonable tube velocities and hence reasonable heat-transfer coefficients. This need for small-diameter tubing and very long flow lengths, which necessitates a number of intermediate water boxes, makes the multistage flash evaporator generally unsuitable for crystallizing service.

It is possible, of course, to combine the different types of evaporator as well as the various methods of energy conservation into a single system. Thus if a crystallizing solution is being handled but saturation is not reached until after an appreciable amount of solvent has been evaporated, the preliminary concentration can be done in several effects of the natural circulation or film type, followed by forced circulation effects for the crystallizing duty. In a thermocompression evaporator of the steam-jet type, one with a steam-turbine-driven mechanical compressor, or one driven by a gas turbine to which a waste-heat boiler can be added, more low pressure steam can be generated than can be used effectively in the thermocompression section. Vapor can, therefore, be bled at the compressor suction and used to operate a multiple-effect or multistage flash-evaporator section. The latter usually can also serve as a preheater for the feed to the thermocompression section. Such evaporators are most useful when the evaporation requirements are the sole or principal reason for installing the facilities for steam or power production. Because it costs only little more to generate steam at a far higher pressure than can be possibly used directly in the evaporator, the addition of a steam-jet or steam-turbine compressor accomplishes additional evaporation at very little

additional operating cost. Similarly, a gas turbine may be relatively inefficient when generating only power but if also used to generate low pressure steam in a waste heat boiler, it becomes as efficient as a more expensive high pressure boiler and steam turbine system.

Temperature Difference and Heat Transfer

The capacity of a steam-heated evaporator is governed by the amount of heat it can transfer, which is determined in turn by the amount of heating surface area and temperature difference available, and the heat-transfer coefficients achieved (see HEAT EXCHANGE TECHNOLOGY, HEAT TRANSFER). There are common conventions in the industry for these terms but they are not universally used; therefore, care must be exercised in interpreting and applying available data. Heating surface area is almost always taken as the total tubing area on the side in contact with the liquor being evaporated, which is different from normal heat exchanger practice, where it is always the outside area of the tubes. The definition of temperature difference is open to the widest interpretation and is the most frequent cause of misapplication of data. Considering only one effect of an evaporator, it is the difference between saturated steam temperature on one side of the heating surface and the liquor temperature on the other side that is effective for heat transfer. However, for the effect as a whole, with its associated liquor circuits, vapor piping, and entrainment separators, it is the difference between saturated steam temperature at the inlet to the heating element and the condensing temperature of the evolved vapor at the exit of the vapor discharge piping that determines how the evaporator effect reacts with its surroundings. Certain losses in temperature difference are inherent, including boiling point rise, vapor circuit losses (those from friction and acceleration and deceleration heads of steam and vapor flowing into and through the heating element, entrainment separator, and vapor piping) and liquor circuit losses (the difference between the temperature at which heat is absorbed at the heating surface and the temperature at which it is released in the vapor head). Boiling point rise, or boiling point elevation (BPE), is usually the most important of these losses and is the difference between the boiling point of the solution in the evaporator and the boiling point of pure solvent at the same pressure. Because the vapor cannot condense and give up the bulk of its heat content until it has cooled to the boiling point of the pure solvent, BPR represents a loss of available ΔT and is deducted from the overall ΔT before computing heat-transfer coefficients. It is usually estimated from known properties of the solution at the discharge concentration of each effect or each body, but care must be exercised here also, because it is sometimes obtained from direct measurements of vapor temperature made at a point where the vapor may still be somewhat superheated. Furthermore, in the falling-film-type evaporator, eg, the discharge is frequently taken from the circulating pump that feeds the heating surface, with the result that the discharge from the heating surface is at a higher concentration than the discharge from the effect.

Any pressure drop losses of steam in the heater or of vapor in the entrainment separator and vapor piping also reduce condensing temperatures, and therefore represent a loss in ΔT available for heat transfer. These losses become more serious at the lower temperatures, because of the increasing difficulty of handling lower density vapor and the decreasing slope of the vapor pressure curve. In many cases, these vapor circuit losses are tabulated separately and deducted from the overall ΔT before computing heat-transfer coefficients. However, at times these losses are ignored or combined with the BPR loss or the heat-transfer ΔT , which must be borne in mind when interpreting heat-transfer data. Incorporating these losses in the BPR term is especially dangerous because actual chemical BPR changes very little with evaporator load whereas the net ΔT across the heating surface increases in almost direct proportion to load, and vapor circuit losses increase as almost the square of the load.

Losses of ΔT in the liquor circuit may also be substantial but are not always taken into consideration. Only in the forced circulation evaporator are at least part of these losses usually calculated. Because heat is absorbed as sensible heat, a temperature rise results that can be either measured or calculated from known heat input and known circulation rate. If this is done, coefficients are reported on the basis of log mean temperature difference. The submergence or short-circuiting losses are frequently included in the coefficient, eg, when the heater inlet temperature is not measured directly but instead is calculated from saturation temperature at the measured vapor head pressure, plus BPR. In natural circulation evaporators the boiling point at the heating surface is higher than that in the vapor space because of the hydrostatic head of liquid above the heating surface. This loss was frequently allowed for in early data presentations. For LTR or LTV evaporators the course of the liquor temperature as it passes the heating surface cannot readily be measured or calculated and coefficients are based on steam and vapor head temperatures. In general, it is practical to consider these losses only when heat-transfer coefficients can be calculated from theory, as in forced circulation and falling-film evaporators. For the latter, the loss caused by friction and acceleration of vapor generated in the tubes must be considered when calculating coefficients from theory, but may or may not have been included in the coefficients reported in the literature.

When designing a new evaporator or analyzing the operation of an existing evaporator, a temperature distribution table first should be prepared. Table 1 gives the data on a quadruple-effect evaporator containing two natural circulation effects followed by two forced circulation effects. Such a table compares the magnitude of the various losses in ΔT and hence indicates the principal obstructions to capacity. Thus if one of the effect's net ΔT s were substantially higher than the others, improvements in performance of that effect would give the greatest gain in capacity. If these data were taken when the evaporator was operating considerably below design capacity, the last effect vapor-circuit loss would be of special concern. It would increase approximately with the square of capacity and soon would become a much more important obstruction as conditions elsewhere were improved.

Table 1. Temperature Distribution, °C

Effect	I	II	III	IV	Condenser
steam temperature ^a	125.7	112.7	94.2	72.8	47.0
net ΔT	9.6	13.4	11.9	12.9	3.0 ^b
liquor temperature	116.1	99.3	82.3	59.9	
BPR ^c	3.2	4.3	6.9	8.6	
vapor temperature ^a	112.9	95.0	75.4	51.3	
vapor-circuit loss	0.2	0.8	2.6	4.3	
heater inlet			83.1	61.6	30
outlet			86.1	64.5	44
net $\Delta T_{\log \text{ mean}}$			9.5	9.7	8.1 ^d
submergence loss			0.8	1.7	

^a Calculated from measured steam chest and vapor-head pressures.
^b Effective ΔT if a direct contact condenser.
^c From known properties of solution, at measured or calculated concentration.
^d Effective ΔT if a surface condenser.

When used in design, a temperature distribution table serves as the basis for calculating a heat and material balance around the evaporator to determine heat loads in the individual effects, vapor flows, and liquor concentrations, from which vapor-circuit losses or line sizes and BPR estimates can be reconfirmed. The designer does not simply take the net ΔT s from such a table and the calculated or estimated heat-transfer coefficients to arrive at the heating surface needed in each effect. The designer also uses it for optimizing. For instance, the temperature distribution may be altered so that the heating surface areas and heating element designs are identical in two or more effects. In the example of Table 1 the ΔT s might be reduced in the first two effects which are of the natural circulation type and cheaper per unit of heating surface, in order to give greater ΔT s and hence lower surface requirements in the more expensive last two forced circulation effects. The added cost of reducing vapor-circuit losses through use of larger lines and entrainment separators

can be compared with the savings resulting therefrom in increased net ΔT s across the heating surfaces.

Actual heat-transfer coefficients encountered in evaporators cover a wide range, depending on the physical properties of the solution, its fouling characteristics, the type of evaporator employed, the boiling temperature, and the temperature difference. Only in the submerged-tube forced circulation and the falling-film evaporator can heat-transfer coefficients easily be calculated from theory (2). For other types, performance estimates are usually based on earlier experience with the same or a similar liquor. In general, coefficients range from a low of about 175 J (m²·s·K) (100 Btu (h·ft²·°F)) to a high of about 3500 (2000). The lowest coefficients are encountered at low temperatures and low ΔT s in film-type or natural circulation evaporators, or at very high viscosities. The highest coefficients are encountered when employing doubly fluted tubes. These high coefficients are possible only when fouling resistances are very low, eg, when handling seawater that has been properly treated for scale prevention and deaerated.

Vapor–Liquid Separation

The heating surface usually determines the evaporator cost and the vapor head the space requirements. The vapor–liquid separator must have enough horizontal plan area to allow the bulk of the initial entrainment to settle back against the rising flow of vapor and enough height to smooth out variations in vapor velocity and to prevent splashing directly into the vapor outlet. Separators are usually sized on the basis of the Souders-Brown expression:

$$U = K \left(\left(\rho_l - \rho_g \right) / \rho_g \right)^{1/2}$$

where U = vapor velocity, ρ_l = density of liquid, ρ_g = density of gas. For most types of evaporator, the decontamination factor DF (DF = kg of vapor per kg of entrained liquid) decreases as K increases, approximately as follows (5):

DF	K	
	m/s	ft/s
100	0.051	0.167
200	0.033	0.108
500	0.020	0.067
1000	0.015	0.049
2000	0.011	0.036

In general, it is not economically attractive to provide the full degree of entrainment separation desired in the vapor head alone. Instead, the vapor head is sized for a decontamination factor on the order of 200 and reliance is placed on supplementary separators such as shown in Figures 1–3 for removal of residual entrainment. Somewhat lower separator velocities are frequently used in crystallizing evaporators to help reduce the buildup of salts on the walls of the vessel above the liquor level. In the falling-film evaporator, most of the entrainment separation occurs within the tubes and higher velocities are permissible in the vapor head and through the curtain of liquor falling from the ends of the tubes. Here, decontamination factors of 1000 or more can be achieved at K values of 0.15 m/s under favorable conditions, primarily avoidance of long fall distances from tube ends to the liquid pool below the tubes, which generates finer, more easily entrainable droplets.

Heat Removal and Noncondensible Gases

A single- or multiple-effect evaporator does not consume heat; it merely degrades the heat input and means must be provided for removing the waste heat. Heat is usually rejected to a river, wells, or a cooling tower. The most common means of heat rejection is by a barometric condenser in which the vapor from the last effect is condensed by direct, countercurrent contact with water cascading over weirs or trays. The condenser is elevated far enough above grade so that water can drain away to a hot well barometrically against the vacuum in the system. Noncondensable gases in the vapor accumulate at the top of the condenser and are cooled close to the temperature of the incoming water, thereby reducing the amount of water vapor associated with the gases. These gases are removed by either a steam-jet ejector (usually of several stages) or a mechanical vacuum pump. Steam-jet ejectors are the most common but are relatively inefficient. Mechanical vacuum pumps have substantially lower operating costs. Another type of direct contact condenser is the cocurrent or jet condenser, in which the water is delivered through nozzles at a velocity high enough to carry the noncondensable gases out the tailpipe. Although this eliminates the need for a separate vacuum pump, the water flow and pressure requirements are sufficiently higher than those of a countercurrent condenser that overall energy consumption is increased. A surface condenser is much more expensive than a direct-contact condenser and is used only when contaminants in the vapor, eg, SO₂, would pollute the condenser water or when the pure condensate has a substantial value. Surface condensers are designed by conventional shell-and-tube methods. Particular care must be taken to minimize shell-side pressure drop, since even a small pressure drop at the high vacuum usually employed represents a substantial loss in available ΔT . Furthermore, in general, more noncondensible gases are present than in the usual steam-heated exchanger and special precautions must be taken to properly channel the vapor flow past the surface between steam inlet and vent outlet, and to allow for the effects of noncondensible gases on heat-transfer performance.

Noncondensible gases are more prevalent in an evaporator system than in most other steam-heated equipment and must be properly handled to avoid serious impairment of heat transfer or reduction in steam economy (6). These may be present as gases dissolved in the feed or liberated by decomposition reactions in the liquor, eg, by bicarbonate breakdown, air in-leakage, and air dissolved in the condenser water. Although venting in practice is almost always empirical and done in excess, the optimum vent rate usually corresponds to a noncondensable gas content of 5–10 vol % in the vents. Vents from heating elements operating under pressure may be released to the atmosphere, whereas vents from vacuum effects are passed through the condenser to remove as much of the associated water vapor as possible ahead of the vacuum pump. Vents are frequently cascaded from one effect to the next on their way to the condenser. Such cascading results in increased concentrations of noncondensible gases in the later effects, and its only potential benefits are the recovery of heat from an upstream vent that has accidentally been opened too wide or to provide sufficient velocity to sweep noncondensibles to the vent. In general, noncondensible gases have little effect on heat-transfer performance provided that they are properly channeled through the evaporator heating elements. This requires a positive vapor flow path from steam inlet to vent outlet, with no pockets of low velocity where noncondensable gases can be trapped, and no low resistance channels that can bypass steam directly from inlet to outlet.

Other Evaporative Methods

Solar evaporation, the first evaporative method developed, is in widespread use, primarily for production of salt from seawater (see CHEMICALS FROM BRINE). Evaporation rates vary widely with climate and can be predicted from weather data and the properties of the solution being evaporated (7). Evaporation rates of pure water are measured experimentally and are published by the National Weather Service. These are determined in small pans and must be reduced about 30% to yield rates experienced in large ponds or reservoirs. Rates must be reduced even further if the material has an appreciable BPR. For seawater saturated with salt, the resultant rate is only about half that of fresh water in small pans. This is of critical importance with substantial rainfall, because a net positive rate based on water in small pans may become a negative rate in large solar ponds containing brine. In solar salt plants the yield almost never approaches the yield expected from the evaporation accomplished. The principal cause is seepage of partially concentrated brine through the pond bottoms and dikes. This seepage has been the principal impediment to application of solar ponds to other uses, such as concentration of oil field brines and other waste streams. Laws require elimination of seepage to avoid groundwater contamination. The cost of providing the seepage barrier exceeds by far all other costs of solar pond development. Solar evaporation has also been used for the production of potable water from seawater. A barrier is required between the ponds and the atmosphere to admit the solar energy and to condense and collect the evaporated water. This barrier is even

more expensive than a pond liner, making the method generally uneconomical. Yields are usually less than half the evaporation rates achieved in open ponds.

At the opposite extreme of fossil fuel usage are single-effect evaporative systems. Conventional steam-heated single-effect evaporators have been discussed. When the BPR is extremely high (as for manufacture of anhydrous NaOH), the evaporator may be heated by molten salt or other high temperature heat-transfer fluids instead of steam. The LTV-type evaporator is normally used in this service. Wiped-film evaporators also sometimes use these high temperature, low pressure heating media to achieve high ΔT s without need for very heavy heat-transfer wall thicknesses. The simplest single-effect evaporative method brings combustion gases into direct contact with the material being concentrated. A spray dryer can be used in this manner to concentrate liquids, and has been so used for high BPR liquids, such as CaCl_2 . Less expensive in cost and more efficient in fuel consumption is the submerged combustion evaporator. In this case, the burner is immersed in the solution or slurry being concentrated and the combustion gases rise through the liquid to release almost all but the latent heat of the water in the combustion gases. Fuel may be either natural gas or the lighter distillate oils. Such evaporators are inexpensive and well adapted to handling corrosive and severely scaling liquids. They require no heat sink, but since the evolved vapor is mixed with large volumes of combustion gas they make it impractical to achieve much better than single-effect steam economy. These evaporators are impractical for all except very low capacities or the most refractory scale-forming or corrosive liquids.

Reverse osmosis (qv) is being used for some of the work previously only feasible by evaporation, such as concentrating dilute solutions that do not deposit solids on being concentrated. The limit is approximately that of recovering fresh water from seawater, where osmotic pressures to be overcome reach about 3600 kPa (520 psi). The choice here depends primarily on energy source. Reverse osmosis requires prime mechanical energy whereas evaporation systems frequently can make use of degraded energy, such as turbine exhaust steam. Evaporators have found applications at the other end of the spectrum, in concentrating waste streams to higher total solids contents and thus smaller volumes than possible by other methods, or to only solids, under the impetus of "zero liquid discharge" requirements. In some cases, this has required very large and efficient systems (6), but many waste streams are of such small volume that evaporators and solids handling techniques in use near 1900 are again appropriate.

Increasing energy costs have pushed optimum steam economy of evaporators higher and higher. This in turn has aided introduction of falling-film evaporators into more services because the heat-transfer performance does not degrade at the lower temperature differences then available. Higher costs of maintenance labor and demand for better onstream time have promoted use of more corrosion-resistant alloys. As an example, kraft-mill evaporators, which comprise a large share of the total steam-heated evaporation load in the United States, have gone from predominantly five- or six-effect mild steel LTV evaporators to eight- or nine-effect stainless steel falling-film evaporators, with appreciably higher single-unit capacities.

Some use is being made of lower grade heat sources, such as moist air from driers, and the construction of auxiliaries, such as condensers, integral with the evaporator body. A further step is elimination of the conventional condenser-cooling tower-vacuum pump circuit by recirculating last-effect liquor over the equivalent of a cooling tower built as an integral part of the evaporator body.

BIBLIOGRAPHY

"Evaporation" in *ECT* 1st ed., Vol. 5, pp. 927–947, by W. L. Badger, Consulting Engineer; in *ECT* 2nd ed., Vol. 8, pp. 559–580, by F. C. Standiford, W. L. Badger Associates, Inc.; in *ECT* 3rd ed., Vol. 9, pp. 472–493, by F. C. Standiford, W. L. Badger Associates, Inc.

1. J. H. Anderson, *ASME Paper 76-WA/Pwr-5*, ASME, New York, 1976.
2. J. G. Knudsen in D. W. Green, ed., *Chemical Engineers' Handbook*, 6th ed., McGraw-Hill Book Co., New York, 1984, Section 10, pp. 34–38.
3. D. Weimer, H. R. Dolf, and D. A. Austin, *Chem. Eng. Progr.* **76**(11), 70–77 (1980).
4. T. M. Fosberg and W. Sweet, Proceedings of the 53rd International Water Conference, Engineering Society of Western Pennsylvania, Pittsburgh, Pa., 1992.
5. L. Rubin in ref. 2, Section 11, pp. 11–40.
6. F. C. Standiford, *Chem. Eng. Progr.* **75**(7), 59 (1979).
7. J. Ferguson, *Aust. J. Sci. Res.* **V**, 315 (1952).

General References

Reference 2. W. L. Badger and J. T. Banchero, *Introduction to Chemical Engineering*, McGraw-Hill Book Co., New York, 1955, Ch. 5.

R. C. Bennett, *Chem. Eng.* **95**(8), 119–127 (1988).

F. C. Standiford, *Chem. Eng.*, 158 (Dec. 9, 1963).

Energy Conservation Tools for the Process Engineer: Upgrading Existing Evaporators to Reduce Energy Consumption, Report COO 2870-2, National Technical Information Service, U.S. Department of Commerce, Springfield, Va. 22161, 1977.

Evaporator testing and troubleshooting

Equipment Testing Procedures—Evaporators, AIChE, New York, 1978.

F. C. Standiford, *Chem. Eng. Progr.* **58**(11), 80 (1962).

H. H. Newman, *Chem. Eng. Progr.* **64**(7), 33 (1968).

Desalination evaporators

The Office of Saline Water (now the Off. of Water Res. Tech.) of the U.S. Dept. of Interior published almost 1000 reports and was responsible for numerous articles on desalination. The Nuclear Desalination Information Center of Oak Ridge National Laboratory maintains a bibliography of these sources of information, which are abstracted and indexed in ORNL-NDIC-11 and -13.

N. Lior, ed., *Measurement and Control in Water Desalination*, Academic Press, Inc., New York, 1986.

Ferris C. Standiford
Consultant

EXHAUST CONTROL, AUTOMOTIVE

Automobiles and trucks consume large amounts of gasoline, producing commensurately large amounts of gaseous exhaust consisting primarily of carbon dioxide [124-38-9] (qv), water, unburned hydrocarbons (qv), carbon monoxide [630-08-0] (qv), and oxides of nitrogen, NO_x (see GASOLINE AND OTHER MOTOR FUELS). The latter three atmospheric pollutants have been regulated since the 1970s by the U.S. government and more stringently by the State of California (see AIR POLLUTION). Automobile companies have developed fuel metering and exhaust systems using the catalytic converter to meet emission regulations. Carbon dioxide emissions are indirectly controlled by corporate average fuel economy (CAFE) standards for passenger cars and small trucks.

Prior to emission control, passenger car and truck emissions together were the largest contributors to atmospheric pollution in the United States (1). By 1994, when the Tier I exhaust emission standards mandated by the Clean Air Act Amendments of 1990 take effect in the United States, the degree of cleanup for automobiles and small trucks must be 97.4° for hydrocarbon emissions, 96.0° for carbon monoxide, and 90° for oxides of nitrogen as compared to pre-control of exhaust emissions. For areas in the United States that do not reach minimum ambient air standards by the end of the 1990s, the U.S. Congress has set conditional Tier II standards for vehicles to take effect as early as the year 2003. Also, California, because of unusually severe atmospheric pollution conditions, has established the most stringent automobile exhaust regulations in the world to control exhaust emissions to the absolute minimum levels. These California standards, called Low Emission Vehicle (LEV) Standards, start to take effect in 1996. Other states are expected to adopt the California standards.

Key to achieving the degree of exhaust emission control by the early 1990s were the monolithic catalytic converter, the three-way catalyst, and the closed loop system based on the oxygen (qv) sensor (see SENSORS). Initially, the catalytic converter contained an oxidation catalyst to control hydrocarbon (HC) and carbon monoxide emissions. Exhaust gas recirculation was used to control NO_x emissions. Subsequently, the three-way catalyst was developed to control HC, CO, and NO_x in a single catalyst. Over 200 million vehicles have been equipped with the catalytic converter which has been rated among the top 10 engineering breakthroughs of the twentieth century (2). Emission control is achieved without negatively affecting fuel economy or performance. The shift to alternative transportation fuels such as methanol, ethanol, natural gas, liquefied petroleum gas (LPG), and reformulated gasoline in accordance with the Clean Air Act and the National Energy Policy Act of 1992 (see ALCOHOL FUELS; GAS, NATURAL) is expected to produce further modifications to the catalytic converter.

Diesel engine emission control technology is not discussed herein. As of early 1993 only one passenger car manufacturer was marketing diesel fuel cars. Emission control technology for diesel engines, used in some light-duty trucks and in medium- and heavy-duty trucks, is evolving at a rapid pace. This technology includes engine modifications such as high pressure fuel injection, variable valve timing, and intercooled turbochargers. Catalytic after treatment that is quite different from that discussed herein has also been developed.

Emission Regulations and Testing

In the United States, federal regulations require automobile manufacturers to certify that vehicles are in compliance with exhaust emission standards when tested under specific test procedures.

Clean Air Act Amendments. Exhaust emission standards were set by the Clean Air Act Amendments of 1970 requiring automobile manufacturers to control the amount of hydrocarbons, carbon monoxide, and oxides of nitrogen emitted from vehicles. The act was amended in 1990. Table 1 summarizes the emission standards for passenger cars and light trucks between the years 1991 and 2006. In 1994, emissions from vehicles come under more stringent control, and the useful life of the emission control system is extended from five years or 80,000 km to 10 years or 160,000 km. The regulations are set up so that after 80,000 km or five years of actual use, a slightly more relaxed standard is applicable for the next 80,000 km.

Table 1. Federal Exhaust Emission Standards for Conventionally Fueled Passenger Cars and Light Trucks^a, g/km

Vehicle and year	Fleet, °	NMHC ^b	CO	NO _x	Particulates
<i>Light-duty (0–2,700 kg gross vehicle weight rating) vehicles</i>					
passenger cars					
1991–1993	100	0.25 ^c	2.1	0.6	0.12
1994	40	0.16 ^d	2.1	0.2	0.05
1994 ^e	40	0.19	2.6	0.4	0.06
1995	80	0.10 ^d	2.1	0.2	0.05
1995 ^e	80	0.19	2.6	0.4	0.06
1996–2003	100	0.10 ^d	2.1	0.2	0.05
1996–2003 ^e	100	0.19	2.6	0.4	0.06
2004–2006 ^{e,f}		0.08	1.1	0.1	0.06
trucks, loaded vehicle weight <1700 kg					
1991–1993	100	0.50 ^c	6.2	0.7	0.08
1994	40	0.16 ^d	2.1	0.2	0.08
1994 ^e	40	0.19	2.6	0.4	
1995	80	0.16 ^d	2.1	0.2	0.05 ^g
1995 ^e	80	0.19	2.6	0.4	0.06 ^g
1996–2003	100	0.16 ^d	2.1	0.2	0.05 ^{h,1}
1996–2003 ^e	100	0.19	2.6	0.4	0.06 ^{h,1}
2004–2006 ^{e,f}		0.08	1.1	0.1	0.6
trucks, loaded vehicle weight 1700–2600 kg					
1991–1993 ¹	100	0.50 ^c	6.2	1.1	0.08
1994	40	0.20 ^d	2.7	0.4	
1994 ^e	40	0.25	3.4	0.60	0.08
1995	80	0.20 ^d	2.7	0.4	0.05 ^g

1995 ^e	80	0.25	3.4	0.60	0.06 ^{g,h}
1996–2003	100	0.20 ^d	2.7	0.4	0.05 ^{h,i}
1996–2003 ^e	100	0.25	3.4	0.60	0.06
<i>Light-duty (2,700–3,900 kg gross vehicle weight rating) trucks</i>					
trucks, 1,700–2,600 test weight ^k					
1991–1993 ^l	100	0.50 ^c	6.2	1.1	0.08
1994 ^l	100	0.50 ^c	6.2	1.1	0.08
1995 ^l	100	0.50 ^c	6.2	1.1	0.08
1996	50	0.20	2.7	0.4	
1996 ^l	50	0.28 ^d	4.0	0.61	0.06
1997–2003	100	0.20 ^d	2.7	0.4	
1997–2003 ^e	100	0.28	4.0	0.61	0.06
trucks, >2,600 kg test weight ^k					
1991–1993 ^k	100	0.50 ^c	6.2	1.1	0.08
1994 ^l	100	0.50 ^c	6.2	1.1	0.08
1995	100	0.50 ^c	6.2	1.1	0.08
1996 ^l	50	0.24 ^d	3.1	0.7	
1996 ^l	50	0.35	4.5	0.95	0.07
1997–2003	100	0.24 ^d	3.1	0.7	
1997–2003 ^e	100	0.35	4.5	0.95	0.07

^a The useful life of the emissions control system is expected to be five years or 80,000 km, unless otherwise noted.

^b NMHC = nonmethane-reactive hydrocarbons.

^c Total hydrocarbons.

^d A set of intermediate in-use standards also applies during the phase-in period, 1994–1997 for passenger cars and small light-duty trucks, and 1996–1998 for larger light-duty trucks.

^e Useful life is 10 years or 160,000 km; in-use compliance is seven years or 120,000 km.

^f After an EPA OTA study due June 1, 1997, the EPA shall by December 31, 1999, set standards more stringent than 1996 standards for passenger cars and certain light trucks, effective after January 1, 2003, but not later than model year 2006.

^g Corresponds to 40% of fleet for particulate.

^h Corresponds to 80% of fleet for particulate.

ⁱ 100% of fleet in 1997 and thereafter.

^l Useful life is 11 years or 192,000 km; in-use compliance is seven years or 144,000 km starting 1994.

^k Test weight = (gross vehicle weight rating + curb weight) / 2.

Test Procedure. To comply with emission standards, representative vehicles must be run for 80,000 km (Appendix IV of the Federal Test Procedure (FTP)) (3). The first 6,400 km are considered a break-in portion. Exhaust emissions are measured each 8,000 km between approximately 6,400 and 80,000 km of accumulation and a deterioration factor (DF) of emissions is calculated. A DF of 1.15 for HC indicates that HC emissions increased by 15% between 6,400 and 80,000 km, and were within the 80,000 km standard. This DF is applied to the 6,400 km emission test data points for all other model variations of the family of vehicles represented by the 80,000 km durability car.

The test for evaluating individual vehicle emissions, the FTP (4), specifies that a test vehicle be stored in an area where the ambient temperature is between 20 and 29°C for at least 12 hours immediately prior to the emission test. Then, the vehicle is placed on a chassis dynamometer which is calibrated for the vehicle weight and road load. The vehicle is started and driven for 41 min on a prescribed cycle of accelerations, cruises, decelerations, idles, a 10-min shutdown (called the hot soak), and a period of rerun.

During the FTP, all exhaust passes through a constant volume sampling system (CVS) which dilutes the exhaust with air so that the total flow of air plus exhaust is constant. A sample of the diluted gas is metered to three gas sample collection bags sequentially. The first, or cold start bag, contains the gas sample from the first 505 seconds of the test. The second bag is the gas sample of the hot transient portion of the test, between 506 and 1372 seconds, after which time the ignition is turned off for 10 minutes. The third bag gas sample is taken after the 10 min hot soak, from the point of ignition of the hot restart, for 505 seconds. Each bag is then measured for hydrocarbons, carbon monoxide, and oxides of nitrogen using, respectively, a flame-ionization detector (FID) HC analyzer, an infrared CO analyzer, and a chemiluminescent NO_x analyzer. A CO₂ analyzer (NDIR) is used to calculate dilution. The concentration measurements are converted into grams of each emission per unit of distance. The FTP prescribes weighting factors for each phase to give a composite value for the total test.

Typical HC and CO emissions from a vehicle undergoing the FTP are shown in Figure 1a. Figure 1b shows the inlet gas temperature to a typical underfloor catalytic converter. The HC and CO are very high during the first 100 seconds after the engine is started, and drop off precipitously after the engine is up to running temperature. From 75 to 85% of the HC and CO emissions pass through the catalytic converter during the time the converter is being heated by the hot exhaust gases (5). Very little of these gases pass unconverted through a hot catalyst of a properly calibrated system.

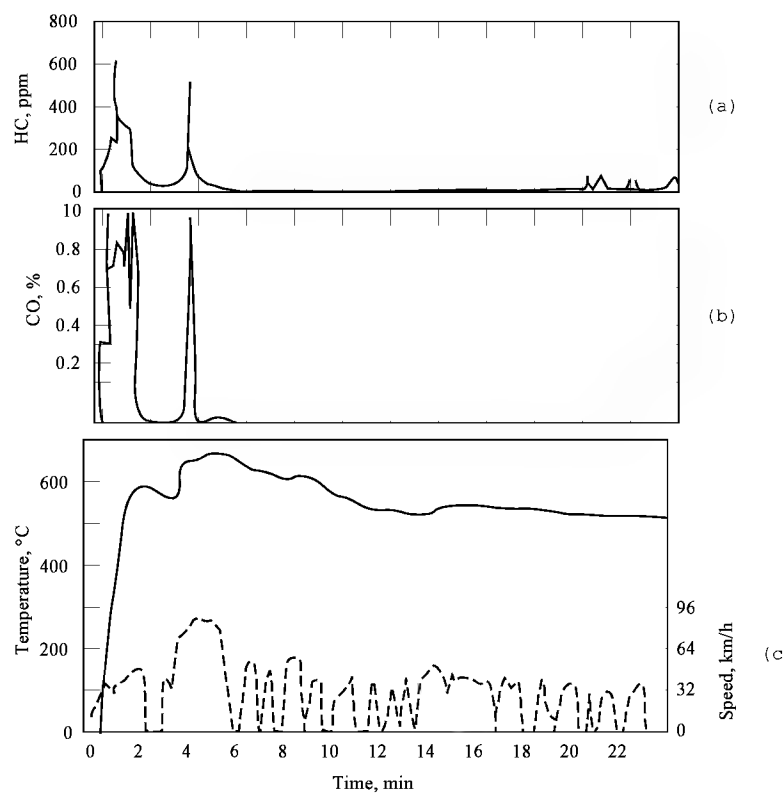


Fig. 1. CO and hydrocarbon tailpipe emissions. Data from a test vehicle during a test cycle where the catalyst was mounted ~ 1.2 m from the exhaust part of the engine: (a) hydrocarbon and (b) CO tailpipe emissions; and (c), (—) catalyst temperature and (---) speed. As can be seen, the principal CO and hydrocarbon emissions occur during the cold start for this vehicle, ie, during the catalyst warmup period. When hot, the catalyst is very effective. In practice, one can expect between 60 and 90% of the engine CO and hydrocarbon emissions, as measured over the whole test cycle, to be removed by the catalyst after 50,000 miles of use (6).

Exhaust Gas Composition

The exhaust composition from gasoline–air combustion is dependent on many factors. Total combustion in the engine is not possible even when excess oxygen is present (6). Formation of the air–fuel mixture as well as the design of the combustion chamber influence the combustion process, as do engine power and ignition system timing. However, for emission control, the main factor affecting the composition of the exhaust gas is the air–fuel mixture or ratio. The composition of exhaust varies according to the air–fuel ratio as shown in Figure 2 (7) for a standard gasoline fuel where the hydrogen to carbon ratio (H/C) is 1.86. The exhaust gas composition changes as the H/C ratio changes (9), as shown in Figure 3 (8,10,11).

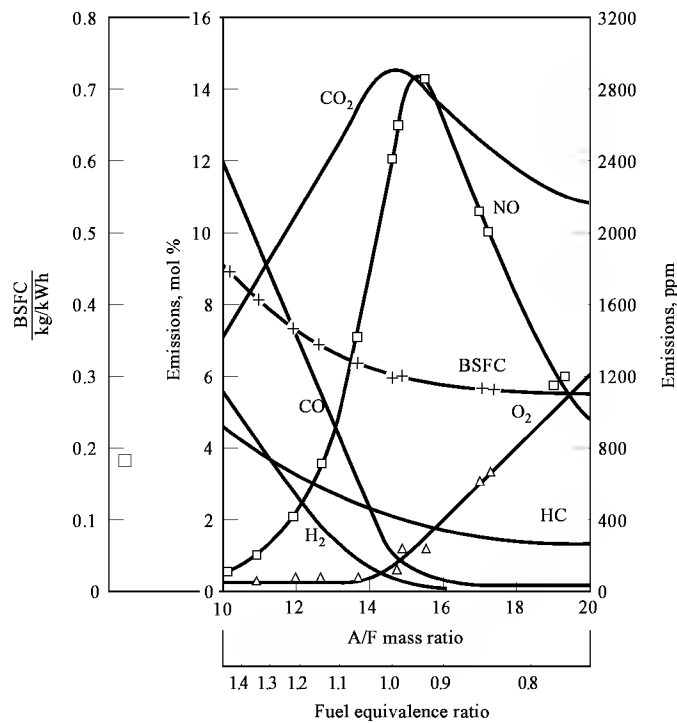


Fig. 2. Effect of mixture strength on exhaust gas composition (dry basis) and brake specific fuel consumption (BSFC) for an unsupercharged automotive-type engine using indolene fuel, H/C = 1.86, where the ignition is tuned to achieve maximum best torque (MBT), the brake mean effective pressure (BNEP) is 386 kPa at 1200 rpm (7,8).

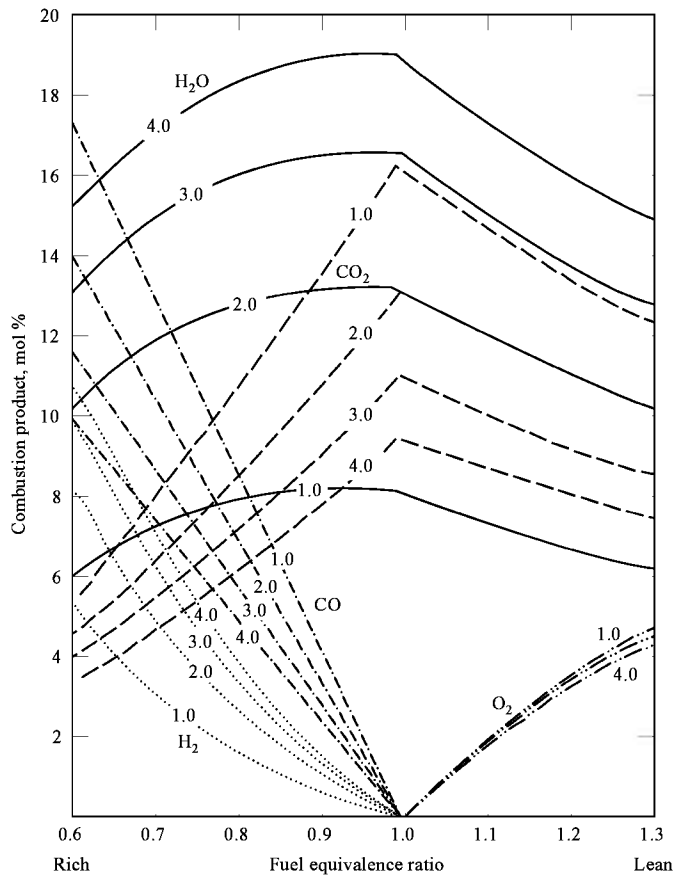


Fig. 3. Theoretical mole percent of the principal combustion products of hydrocarbon fuels for fuel hydrogen:carbon ratios from 1, eg, C H , to 4, eg, CH₄, wet basis, where (—) represents H₂O; (---) CO₂; (-·-·-) CO; (- - - -) O₂; and (·····) H₂ (8,10,11).

Unburned hydrocarbons in the exhaust originate primarily from crevices in the combustion chamber, such as gaps between the piston and cylinder wall, where the combustion flame cannot burn. The composition of unburned hydrocarbons is dictated primarily by the composition of the fuel (12). Carbon monoxide results from areas of insufficient oxygen. Oxides of nitrogen are produced in the high temperature zones during combustion by the reaction of nitrogen molecules and oxygen atoms thermally produced from oxygen and oxygen-containing species, according to the Zeldovich mechanism (6,13).

Hydrocarbons and carbon monoxide emissions can be minimized by lean air /fuel mixtures (Fig. 2), but lean air /fuel mixtures maximize NO_x emissions. Very lean mixtures (>20 air /fuel) result in reduced CO and NO_x, but in increased HC emissions owing to unstable combustion. The turning point is known as the lean limit. Improvements in lean-burn engines extend the lean limit. Rich mixtures, which contain excess fuel and insufficient air, produce high HC and CO concentrations in the exhaust. Very rich mixtures are typically used for small air-cooled engines, needed because of the cooling effect of the gasoline as it vaporizes in the cylinder, where CO exhaust concentrations are 4 to 5% or more.

The best power is achieved slightly rich of stoichiometric air /fuel; the best fuel economy is achieved slightly lean of stoichiometric mixture.

Over 150 hydrogen and carbon species are present in the exhaust of a gasoline fueled engine (14–18) including methane, various paraffins, olefins, aldehydes (qv), aromatics, and polycyclic hydrocarbons as well as unburned gasoline. Exhaust gas also contains sulfur dioxide from the combustion of sulfur contained in the gasoline. The average U.S. gasoline sulfur content is about 300 ppm, which results in about 20 ppm SO₂ in the exhaust. Gasoline sulfur has ranged between 50 and >1000 ppm in the United States. Reformulated gasolines are expected to have lower sulfur because of the greater processing by the petroleum (qv) refinery (11,12). Additionally, the State of California has specified a Phase II gasoline having a sulfur content between 30–40 ppm for the LEV program.

Exhaust gas also contains small amounts of hydrogen cyanide and ammonia depending on the air /fuel ratio.

Emission Control System

A typical 1993 model year automobile emission control system containing a multipoint fuel injection fuel metering system, which meters the fuel in response to a measured amount of air, is shown in Figure 4. Incorporated into the exhaust stream prior to the catalytic converter is an oxygen sensor which indicates whether the exhaust air /fuel mixture is rich or lean of the stoichiometric point (defined as neither air nor fuel in excess). The catalytic converter is located in the exhaust line leading from the exhaust manifold, upstream of the acoustic muffler. The signal generated by the oxygen sensor is sent to the computer controller as is a signal from an air flow measurement device. The computer controller regulates the fuel metered by the fuel injection system in response to the air measurement signal. Air flow varies in response to the throttle position and load (inlet vacuum). The oxygen sensor quickly detects any change in oxygen concentration in the exhaust and the controller adjusts for this change. Thus the air /fuel ratio is constantly being adjusted back and forth slightly rich and slightly lean of the stoichiometric mixture. The three-way conversion (TWC) catalyst therefore receives exhaust gas that reflects this constant change back and forth in air /fuel mixture and is designed to operate under those conditions to convert NO_x by reduction and HC and CO by oxidation of at least 80 to 90%.

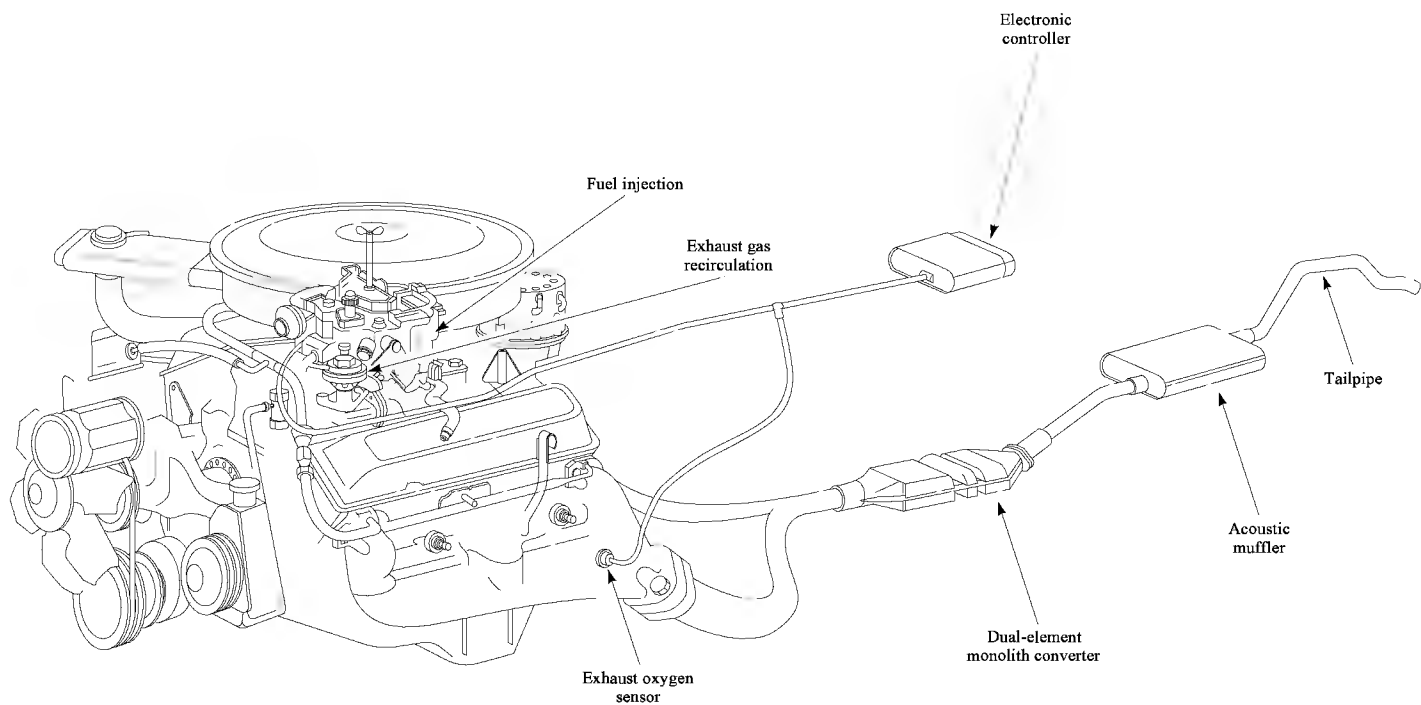


Fig. 4. Closed-loop dual-catalyst system for emissions control using dual element mono-lith converter, which is three-way and oxidizing.

Catalytic Converter. The converter consists of a catalytic unit contained in a metal canister which surrounds the fragile ceramic catalytic unit with a steel shell. The converter shell assembly is usually made from an iron-chrome Series 409 muffler-grade stainless steel that is resistant to internal and external oxidative corrosion. In between the steel shell and the exterior of the catalytic unit is a compliant layer that grips the catalytic unit with sufficient force to prevent movement of the catalytic unit within the canister, and which compensates for the differences in thermal expansion between the catalyst and the metal shell. Several types of compliant layers are used, and all have spring-like properties under compression that provide the necessary gripping force at all exhaust temperatures. Corrugated knitted wire mesh was the first successful compliant layer. As of this writing, a material based on vermiculite, which expands upon application of heat (about 300°C), is used, as is a wire mesh material wrapped several times around the catalytic unit. The compliant layer mounting system has proved to be durable for the life of the vehicle. The converter design, flow, and pressure drop characteristics are described in the literature (19–23).

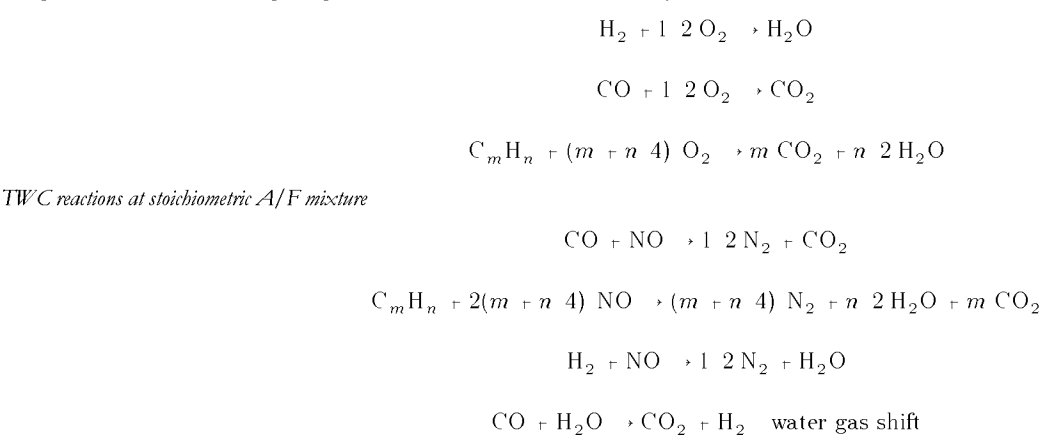
The catalytic unit is designed to provide enough surface area so that all exhaust gases contact the catalyst surfaces as they pass from the engine to the tailpipe (24–27). In order to function quickly after the engine is started, the catalytic unit must rapidly heat up to operating temperature. It therefore must possess good heat-exchange properties to extract the necessary heat from the exhaust gas. Once the minimum catalytic operation temperature is reached, the catalytic unit is designed to maximize transfer of the pollutants from the exhaust gas to the surface of the catalytic unit. Heat transfer and mass transfer are driven by temperature difference and concentration difference, respectively. At operating temperatures above 300 or 350°C, the catalytic reactions are so fast that only the exterior surfaces of the catalyst are utilized for the catalytic function (28).

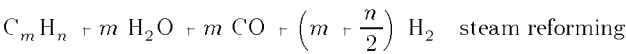
Automobile exhaust catalysts have been developed that maximize the catalyst surface area available to the flowing exhaust gas without incurring excessive pressure drop. Two types have been extensively studied: the monolithic honeycomb type and the pellet type.

Use of the pelletized converter, developed and used by General Motors starting in 1975, has declined since 1980. The advantage of the pelletized converter, which consists of a packed bed of small spherical beads about 3 mm in diameter, is that the pellets were less costly to manufacture than the monolithic honeycomb. Disadvantages were the pelletized converter had 2 to 3 times more weight and volume, took longer to heat up, and was more susceptible to attrition and loss of catalyst in use. The monolithic honeycomb can be mounted in any orientation, whereas the pelletized converter had to be downflow. Additionally, the pressure drop of the monolithic honeycomb is one-half to one-quarter that of a similar function pelletized converter.

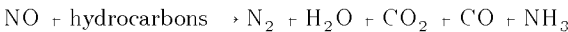
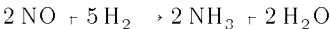
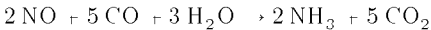
Pelletized converters are used by General Motors for heavy-duty gasoline fueled trucks because the pellets have better high temperature physical characteristics and the converter has been redesigned to minimize pressure drop for this application (19). However, these are expected to be replaced by monolithic converters. In Japan, taxicabs have used pelletized converters and LPG (propane) fuel since the 1960s. The catalyst in these converters is changed once per year. It is thought easier to change pellets than to change a monolithic honeycomb catalyst. Although this practice of changing catalyst remains, the monolithic honeycomb converter remains active and fully functional for the life of the vehicle, especially when operated on a clean fuel such as LPG.

Catalytic Unit. The catalytic unit consists of an activated coating layer spread uniformly on a monolithic substrate. The catalyst predominantly used in the United States and Canada is known as the three-way conversion (TWC) catalyst, because it destroys all three types of regulated pollutants: HC, CO, and NO_x. Between 1975 and the early 1980s, an oxidation catalyst was used. Its use declined with the development of the TWC catalyst. The TWC catalytic efficiency is shown in Figure 5. At temperatures of >300°C a TWC destroys HC, CO, and NO_x effectively when the air-fuel mixture is close to stoichiometric, as shown. Other conventions for describing the air-fuel mixture are use of the Greek letter lambda (λ) where λ = 1.0 is the stoichiometric mixture, REDOX ratio, and equivalence ratio. When the air-fuel mixture is near the stoichiometric point, an optimum value of conversion for all three components is achieved. The principal chemical reactions of a TWC catalyst are

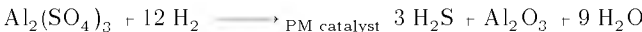
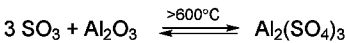
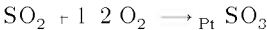
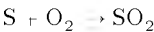




Other NO-reduction reactions



Fuel sulfur reactions



where PM is precious metal (6,29).

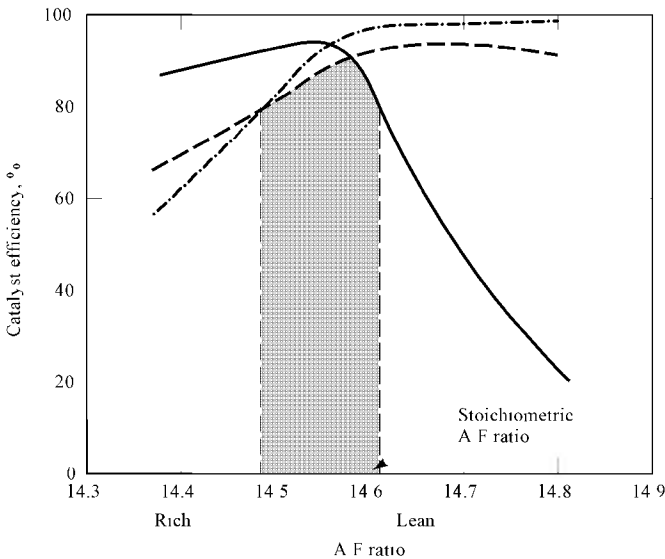


Fig. 5. Conversion of (—) NO_x, (- · - ·) CO, and (---) hydrocarbons for a TWC as a function of the air / fuel (A / F) ratio. The shaded area shows the A / F ratio window where the TWC catalysts function at 80% efficiency at 400°C.

Activated Coating. The activated coating layer of a TWC is applied to the high geometric surface area monolithic honeycomb body or substrate. Precious metals are dispersed within this catalytic layer, which contains particles as small as 10 μm. The thickness of the layer is between 25 and 100 μm. The honeycomb substrate typically has a geometric surface area of 27 cm² / cm³ so that a typical vehicle catalyst volume of 1.6 L would contain 4.6 m² of geometric surface area. The activated coating layer increases the total surface to 7000 m², ie, more than three orders of magnitude. The total surface area consists of the combined internal and external surface of all the minute particles contained in the activated coating layer. Total surface area is measured by the classic Brunauer-Emmet-Teller (BET) technique, which involves N₂ adsorption at low temperature (30).

The small (10 μm) coating particles are typically aluminum oxide [1344-28-1], Al₂O₃. These particles can have BET surface areas of 100 to 300 m² / g. The thermal and physical properties of alumina crystalline phases vary according to the starting phase (aluminum hydroxide or hydrate) and thermal treatment (see ALUMINUM COMPOUNDS, ALUMINUM OXIDE).

Alumina is used because it is relatively inert and provides the high surface area needed to efficiently disperse the expensive active catalytic components. However, no one alumina phase possesses the thermal, physical, and chemical properties ideal for the perfect activated coating layer. A great deal of research has been carried out in search of modifications that can make one or more of the alumina crystalline phases more suitable. For instance, components such as ceria, baria, lanthana, or zirconia are added to enhance the thermal characteristics of the alumina. Figure 6 shows the thermal performance of an alumina-activated coating material.

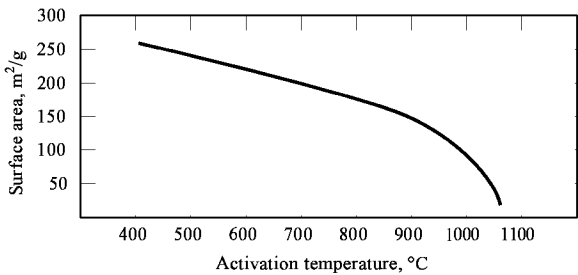


Fig. 6. Surface area stability of Pural SB. The activation time was 3 h.

An unstabilized high surface area alumina sinters severely upon exposure to temperatures over 900°C. Sintering is a process by which the small internal pores in the particles coalesce and lose large fractions of the total surface area. This process is to be avoided because it occludes some of the precious metal catalyst sites. The network of small pores and passages for gas transfer collapses and restricts free gas exchange into and out of the activated catalyst layer resulting in thermal deactivation of the catalyst.

The activated coating layer must possess two additional properties. It must adhere tenaciously to the monolithic honeycomb surface under conditions of rapid thermal changes, high flow, and moisture condensation, evaporation, or freezing. It must have an open porous structure to permit easy gas passage into the coating layer and back into the main exhaust stream. It must maintain this porous structure even after exposure to temperatures exceeding 900°C.

Precious Metal Catalysts. Precious metals are deposited throughout the TWC-activated coating layer. Rhodium plays an important role in the reduction of NO_x, and is combined with platinum and or palladium for the oxidation of HC and CO. Only a small amount of these expensive materials is used (31) (see PLATINUM GROUP METALS). The metals are dispersed on the high surface area particles as precious metal solutions, and then reduced to small metal crystals by various techniques. Catalytic reactions occur on the precious metal surfaces. Whereas metal within the crystal cannot directly participate in the catalytic process, it can play a role when surface metal oxides are influenced through strong metal to support reactions (SMSI) (32,33). Some exhaust gas reactions, for instance the oxidation of alkanes, require larger Pt crystals than other reactions, such as the oxidation of CO (34).

The small precious metal crystals can exist as metal crystallites or as metal oxides, both of which are catalytic (31). Rodium oxide has a tendency to react with alumina to form a solid solution (35). To minimize this reaction, zirconia is used with the alumina (36). Publications regarding the TWC function of precious metals abound (37–42).

Catalytic Support Body Monolithic Honeycomb Unit. The terms substrate and brick are also used to describe the high geometric surface area material upon which the active coating material is placed. Monolithic honeycomb catalytic support material comes in both ceramic and metallic form. Both are used in automobile catalysts and each possesses unique properties. A common property is a high geometric surface area which is inert and does not react with the catalytic layer.

Ceramic. The ceramic substrate is made from a mixture of silicon dioxide, talc, and kaolin to make the compound cordierite [12182-53-5]. Cordierite possesses a very low coefficient of thermal expansion and is thermal-shock resistant. The manufacturing process involves extruding the starting mixture (which is mixed with water and kneaded into a sort of dough) through a complex die to form the honeycomb structure. The extruded piece is dried and fired in a kiln to form the cordierite. The outside or circumferential dimension is formed by the die, and the length is cut later with a ceramic saw.

The extrusion of the honeycomb was a significant advance aiding the adoption of the monolithic honeycomb catalyst by the automobile industry as the preferred means of emission control (43). Coupled with this was the development of the mounting system (44) securing the fragile ceramic substrate-based catalytic unit in a protective canister. Some of the physical properties considered when selecting a substrate for a catalytic converter are given in Table 2. The most common cell structure used is 62 cells cm² with a 0.152 mm thick wall. This cell density is about optimal with respect to pressure drop, geometric surface area, physical strength, and general ruggedness (24–27).

Table 2. Physical Properties of Catalytic Converter Substrates

Parameter	Cordierite value		
Cell structure			
wall thickness, mm	0.152	0.203	0.304
cells cm ²	62	46	46
pressure drop, °o	100	90	120
geometric surface area, cm ² cm ³	27	24	22
isostatic strength, ^a	>103	>103	>206
bulk density, g cm ³	0.4	0.4	0.5
cross-sectional			

^a To convert N m²(= Pa) to kgf cm², multiply by 1.02 × 10⁻⁵.

Metallic. Metallic substrates are also used as a support for the activated coating layer. A class of metal alloys containing Fe, Cr, and Al, when stabilized with Y or Ce, have excellent oxidation resistance at the extreme temperatures found in automobile exhaust (45). Melting temperatures are about in the same range as that of the ceramic cordierite (see HIGH TEMPERATURE ALLOYS). A feature of these alloys is that upon heating in air, an aluminum oxide surface layer is formed. The yttria or ceria is present to stabilize the aluminum oxide surface and resist the spalling of the alumina which would otherwise occur. The ceria-stabilized alloy develops a whisker-like surface. The catalytic activated coating layer has been found to adhere very well to these surfaces.

The metallic substrate is designed to provide a cell density similar to that of the ceramic counterpart, and can be constructed in versions of 46, 62, or 93 cells cm². However, the wall thickness can be thinner, ie, 0.05 to 0.15 mm thick (46). Additionally, 0.05 mm thick wall material can be constructed in versions of 124 to 248 cells cm². The unique advantage that a metal substrate has over its ceramic counterpart is that either the same geometric surface can be made into a smaller volume, or that for the same volume and geometric surface area, there is a lower pressure drop (47). The 124 to 248 cells m² metal versions have an additional potential performance advantage over ceramic because of large increases in geometric surface area. Metallic substrate catalysts are used by Porsche for high performance vehicles and by Chrysler for the Viper model (45,48). In both of these cases, the lower exhaust gas pressure drop results in increased horse-power and performance. In the future, metallic supports may be used for small catalysts located close to the exhaust manifold in order to heat up faster in emission control systems calibrated to meet the strict California Low Emission Standards.

Mass Transfer. Exhaust gas catalytic treatment depends on the efficient contact of the exhaust gas and the catalyst. During the initial seconds after start of the engine, hot gases from the exhaust valve of the engine pass through the exhaust manifold and encounter the catalytic converter. Turbulent flow conditions (Reynolds numbers above 2000) exist in response to the exhaust stroke of each cylinder (about 6 to 25 times per second) times the number of cylinders. However, laminar flow conditions are reached a short (~0.6 cm) distance after entering the cell passages of the honeycomb (5,49–52).

The process of catalyst heating and initiation of the catalytic function is shown in Figure 7, where there are three or four distinct regions. Depending on the location of the catalytic unit in the exhaust system and the thermal mass present prior to the catalytic unit, it can take from 30 to 120 seconds for the catalyst unit to reach the catalyst ignition temperature of approximately 250–300°C (Region I). At temperatures above this ignition point, the activity of the catalyst increases rapidly with temperature (Region II). Some catalyst reaches a point where the sharp rise with temperature abruptly takes on a mild positive slope (Region III). Then, a point is reached at which catalytic performance improves only slightly to an increase in temperature (Region IV).

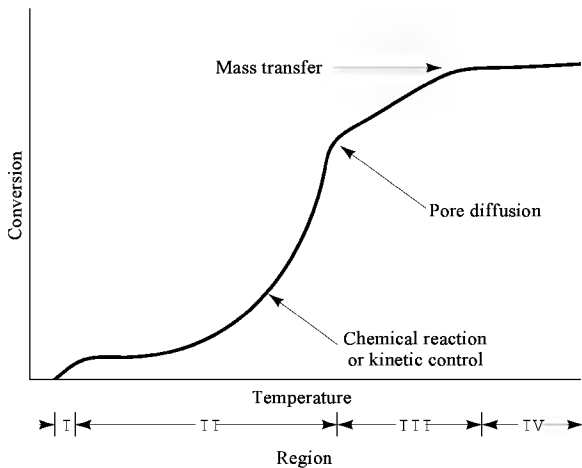


Fig. 7. Conversion vs temperature.

In Region I (below the ignition point) there is no catalytic activity. Within Region II, the specific activity of the catalyst is the rate-limiting step; this is called the kinetically controlled region. Highly active catalysts have a lower ignition point and exhibit large increases in catalytic performance associated with small increases in temperature. The behavior of a catalyst in Region II is important to selecting an auto emission catalyst because catalyst light-off is a prime factor in achieving adequate emission control early in the FTP test.

Region III is not present in all catalysts. It depends on the porous structure design of the catalyst and is often only found in a used catalyst. If present, the rate-limiting step is called pore diffusion control. The volume of the catalytic unit dictates the catalyst performance in Regions III and IV. Catalysts clogged with masking poisons such as lube oil ash would exhibit Region III behavior. Region IV is a mass-transfer limited region, ie, catalytic reactions occur so fast that the rate-limiting factor is getting the reactants to the surface of the catalyst. Mathematical models describing the entire process are found in several references (5,49–52).

Catalyst Function. Automobile exhaust catalysts are perfect examples of materials that accelerate a chemical reaction but are not consumed. Reactions are completed on the catalyst surface and the products leave. Thus the catalyst performs its function over and over again. The catalyst also permits reactions to occur at considerably lower temperatures. For instance, CO reacts with oxygen above 700°C at a substantial rate. An automobile exhaust catalyst enables the reaction to occur at a temperature of about 250°C and at a much faster rate and in a smaller reactor volume. This is also the case for the combustion of hydrocarbons.

Concerning the reduction of NO_x, automobile three-way catalysts exhibit a property called selectivity. Catalyst selectivity occurs when several reactions are thermodynamically possible but one reaction proceeds at a faster rate than another. In the case of a TWC catalyst, CO, HC, and H₂ are all potential reductants of NO. On the other hand, O₂ is present, which oxidizes the CO, HC, and H₂. If these oxidation reactions are too rapid, no reductant is available to convert NO. Using modern TWC catalysts, however, NO reduction is fast enough that it is substantially completed before the reductants are consumed by O₂.

Two classes of metals have been examined for potential use as catalytic materials for automobile exhaust control. These consist of some of the transitional base metal series, for instance, cobalt, copper, chromium, nickel, manganese, and vanadium; and the precious metal series consisting of platinum [7440-06-4], Pt; palladium [7440-05-3], Pd; rhodium [7440-16-6], Rh; iridium, [7439-88-5], Ir; and ruthenium [7440-18-8], Ru. Specific catalyst activities are shown in Table 3.

Table 3. Specific Reaction Rates

Catalyst	CO + O ₂ ^a		C ₂ H ₄ + O ₂ ^b	
	Temperature, °C	R ^c	Temperature, °C	R ^c
Pd (wire)	300	500	300	100
Pt (wire)	300	100	300	10
Au (wire)	300	15	300	0.03
Co ₃ O ₄	200	25 + 5	300	0.33
CuO	200	11	300	0.6
CuCr ₂ O ₄	200	5	300	0.8
LaCoO ₃	200	2.3	400	0.53
BaCoO ₃	300	5.3	400	0.1
SnO ₂	300	5.2	400	0.4
MnO ₂	300	3.4	300	0.04
LaMnO ₃	300	2	400	0.3
La _{0.5} Sr _{0.5} MnO ₃	300	1.2	400	0.26
La _{0.5} Pb _{0.5} MnO ₃	300	0.5	400	<0.05
Fe ₂ O ₃	300	0.4	400	0.06
FeCr ₂ O ₄	300	0.33		
Cr ₂ O ₃	300	0.03	300	0.006
NiO	300	1.5		
ZrO ₂	300	0.013	400	0.002

^a 1° o O₂, 1° o CO, 0° o H₂O.

^b 1° o O₂, 0.1° o C₂H₂, ~0.1% H₂O.

^c Units in terms of CO₂ formation mL (min·m²).

The precious metals possess much higher specific catalytic activity than do the base metals. In addition, base metal catalysts sinter upon exposure to the exhaust gas temperatures found in engine exhaust, thereby losing the catalytic performance needed for low temperature operation. Also, the base metals deactivate because of reactions with sulfur compounds at the low temperature end of auto exhaust. As a result, a base metal automobile exhaust

catalyst would need to be considerably larger than a precious metal one and, even if a large bed were used, it would not heat up quickly enough to achieve the catalytic performance demanded of the emission control systems (6).

Catalyst function in an exhaust gas stream can be understood by examining the catalyst performance curves shown in Figures 5 and 8. Figure 8 shows catalyst function at a specific set of flow conditions as the catalyst inlet temperature is slowly increased. This test, conducted in a laboratory reactor, uses a gaseous mix to simulate the composition of automobile exhaust. As the temperature increases from ambient to about 200°C, there is no apparent action on the part of the catalyst to consume any of the reactants. The carbon monoxide present strongly chemisorbs on the surface of the catalyst, and prevents oxygen access. As the inlet gas temperature approaches 200°C, the CO bonds to the metal surface are relaxed. Oxygen molecules are now able to chemisorb, and catalytic ignition occurs.

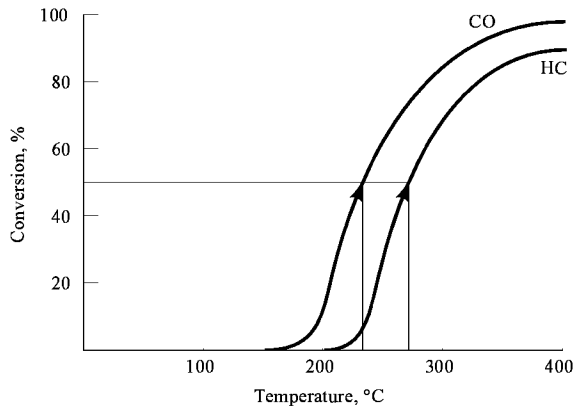


Fig. 8. Effect of temperature on catalyst performance using synthetic exhaust: 0.45% CO, 12% CO₂, 500 ppm NO, 0.15% H₂, 3% O₂, 200 ppm C₃H₈, 10% H₂O, and remainder N₂. The *T*₅₀ for CO and HC are 230°C and 270°C, respectively. See text.

Beyond the catalytic ignition point there is a rapid increase in catalytic performance with small increases in temperature. A measure of catalyst performance has been the temperature at which 50% conversion of reactant is achieved. For carbon monoxide this is often referred to as *T*₅₀ CO. The catalyst light-off property is important for exhaust emission control because the catalyst light-off must occur reliably every time the engine is started, even after extreme in-use engine operating conditions.

A representation of a catalytic reaction is shown in Figure 9. A gaseous pollutant such as CO diffuses from the main gas stream into the porous catalyst matrix. There is a net CO flow from the main flow stream to the catalytic surface because CO is at higher concentration in the exhaust stream, and it is attracted to a precious metal site that is temporarily unoccupied. For the same reasons, oxygen molecules are attracted to the site. The oxygen molecule is chemisorbed onto the precious metal surface where oxygen molecule bond stretching occurs, leading to dissociation to oxygen atoms that bond to the catalyst surface. The rate at which the reaction proceeds depends strongly on the temperature at the catalyst surface. The reaction of carbon monoxide and oxygen proceeds to form CO₂, which desorbs and reenters the exhaust. There is less CO₂ in the flowing exhaust stream and more at the surface so there is a net flow of CO₂ away from the catalyst surface.

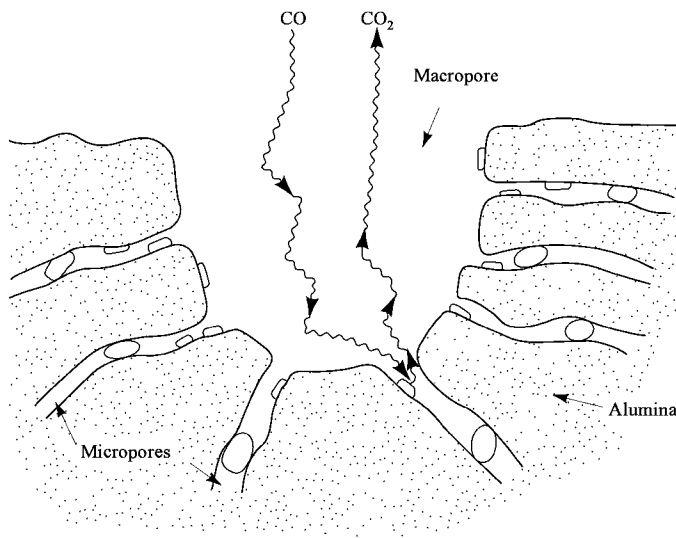


Fig. 9. Catalyst pore and reaction. The CO diffuses into a precious metal site □; reacts with O₂; and leaves as CO₂.

The oxidation of HC and CO must proceed in balance with the reduction of NO_x by CO, HC, or H₂. For the NO_x removal reaction, a reductant is required. First NO is adsorbed on the catalyst surface and dissociates forming N₂ which leaves the surface, but the O atoms remain. CO is required to remove the O atoms to complete the reaction cycle (53).

A TWC catalyst must be able to partition enough CO present in the exhaust for each of these reactions and provide a surface that has preference for NO adsorption. Rhodium has a slight preference for NO adsorption rather than O₂ adsorption; Pt prefers O₂. Rh also does not catalyze the unwanted NH₃ reaction as does Pt, and Rh is more sinter-resistant than Pt (6). However, the concentrations of O₂ and NO have to be balanced for the preferred maximum reduction of NO and oxidation of CO. This occurs at approximately the stoichiometric point with just enough oxidants (O₂ and NO_x) and reductants (CO, HC, and H₂). If the mixture is too rich there is not enough O₂ and no matter how active the catalyst, some CO and HC is not converted. If the mixture is too lean, there is too much O₂ and the NO cannot effectively compete for the catalyst sites (53–58).

In an actual exhaust system controlled by the signal of the oxygen sensor, stoichiometry is never maintained, rather, it cycles periodically rich and lean one to three times per second, ie, one-half of the time there is too much oxygen and one-half of the time there is too little. Incorporation of cerium oxide or other oxygen storage components solves this problem. The ceria adsorbs O₂ that would otherwise escape during the lean half cycle, and during the rich half cycle the CO reacts with the adsorbed O₂ (32,44,59–63). The TWC catalyst effectiveness is dependent on the use of Rh to reduce NO_x and

upon the proper balance of reactants; the oxygen sensor plays the key role in maintaining this proper balance.

Catalyst Durability. Automobile catalysts last for the life of the vehicle and still function well at the time the vehicle is scrapped. However, there is potential for decline in total catalytic performance from exposure to very high temperatures, accumulation of catalyst poisons, or loss of the active layer (29,64–68).

High Temperatures. Exposure to temperatures above 1000°C can cause the active layer surfaces to sinter, collapse, agglomerate, shrink, and possibly exfoliate. Collapse of the surfaces occludes precious metal sites, preventing free passage to these sites. Also, the small precious metal crystals can unite and grow into larger crystals, thereby diminishing the total precious metal surface and thus the total number of active surface sites. If catalyst temperatures exceed 1200°C, the substrate softens and perhaps shrinks. At about 1465°C, the substrate melts and is destroyed. A thermally damaged pore is shown in Figure 10a.

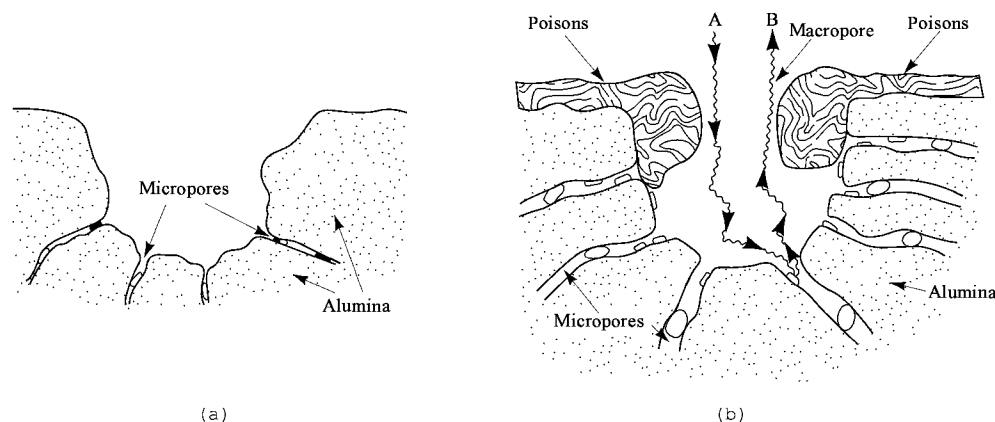


Fig. 10. Catalyst macropores showing □ noble metal sites and (a) narrowed micropores after exposure to high temperatures where ■ represents thermally damaged noble metal sites; and (b) pore mouth plugging from poisons where A, if allowed, diffuses in to be converted to B.

High catalyst temperatures are caused by two types of factors. The first are engine related and can be corrected by calibration. That is, ignition timing and the air–fuel mixture, as well as the engine load, govern the exhaust gas temperature and thus the catalyst inlet gas temperature. The catalyst consumes unburned exhaust components, and most of the corresponding reactions are exothermic, resulting in increased catalyst and exhaust gas temperatures (33). For instance, the combustion of 1% CO would result in a 87°C adiabatic temperature rise of the exhaust gas (33); the combustion of 1000 ppm C₂H₄ would yield a 103°C temperature rise.

The second type is the quality of combustion. A misfire or partial misfire of the air–fuel charge in the combustion chamber causes the unburned air and fuel mixture to pass to the catalyst, where combustion occurs. The temperature rise from combustion in the catalyst is considerable. Two misfiring cylinders in a four-cylinder engine could cause the catalyst to approach the melting temperature of the substrate (1465°C) (69).

There are many reasons for misfire. The principal causes are interruption of ignition energy caused by faulty or wet ignition wiring and either lean or rich noncombustible air–fuel mixtures. Electronic ignition, and the development of very durable ignition wiring, as well as highly improved fuel metering systems, have greatly lessened the occurrence of misfire. Unleaded fuel does not clog, corrode, or wear spark plugs, extending spark plug life. Thus the ignition and fuel metering systems are vastly improved over pre-1975, pre-catalyst equipped vehicles. Catalysts have been designed to resist thermal degradation by the incorporation of base metal oxide stabilizers (70).

Poisons and Inhibitors. Catalyst poisons and inhibitors can come from the fuel, the lube oil, from engine wear and corrosion products, and from air ingestion. There are two types of catalyst poisons: one poisons active sites, the other is a masking agent.

The main catalyst site poison for many years was tetraethyllead [78-00-2], C₄H₉Pb, even after use of unleaded gasoline. Not only is lead a catalyst poison, but automotive source lead is also a health hazard (66). The source of this lead came from manufacture and distribution of leaded and unleaded gasoline in common transport equipment and storage facilities (67). In the early 1990s, so little leaded gasoline was being distributed that Pb contamination was approaching zero (<0.26 ppm L).

The mechanism of poisoning automobile exhaust catalysts has been identified (71). Upon combustion in the cylinder tetraethyllead (TEL) produces lead oxide which would accumulate in the combustion chamber except that ethylene dibromide [106-93-4] or other similar halide compounds were added to the gasoline along with TEL to form volatile lead halide compounds. Thus lead deposits in the cylinder and on the spark plugs are minimized. Volatile lead halides (bromides or chlorides) would then exit the combustion chamber, and such volatile compounds would diffuse to catalyst surfaces by the same mechanisms as do carbon monoxide compounds. When adsorbed on the precious metal catalyst site, lead halide renders the catalytic site inactive.

Lead compounds were not found on the surrounding activated coating layer, rather only associated with the precious metal. The Pt sites are less poisoned by lead than are Pd or Rh sites because the Pt sites are protected by the sulfur in the fuel. Fuel sulfur is converted to SO₂ in the combustion process, and Pt easily oxidizes SO₂ to SO₃ on the catalyst site. The SO₃ reacts with the lead compounds to form PbSO₄, which then moves off the catalyst site so that lead sulfate is not a severe catalyst poison. Neither Pd nor Rh is as active for the SO₂ to SO₃ reaction, and therefore do not enjoy the same protection as Pt.

Sulfur oxides resulting from fuel sulfur combustion often inhibit catalyst performance in Regions II, III, and a portion of Region IV (see Fig. 7) depending on the precious metals employed in the catalyst and on the air–fuel ratio. Monolithic catalysts generally recover performance when lower sulfur gasoline is used so the inhibition is temporary. Pd is more susceptible than Rh or Pt. The last is the most resistant. Pd-containing catalysts located in hotter exhaust stream locations, ie, close to the exhaust manifold, function with little sulfur inhibition (72–74).

Fuel sulfur is also responsible for a phenomena known as storage and release of sulfur compounds. Sulfur oxides (SO₂, SO₃) easily react with ceria, an oxygen storage compound incorporated into most TWC catalysts, and also with alumina. When the air–fuel mixture temporarily goes rich and the catalyst temperature is in a certain range, the stored sulfur is released as H₂S yielding a rotten egg odor to the exhaust. A small amount of nickel oxide incorporated into the TWC removes the H₂S and releases it later as SO₂ (75–79).

Masking agents also deactivate catalysts and are loosely called poisons. Small droplets of lubricating oil, for instance, are emitted from the engine and deposit on the surfaces of the catalyst. Lubricating oil contains a small amount of inorganic elements such as zinc, phosphorus, calcium, and barium (see LUBRICATION AND LUBRICANTS). When the organic fractions of lube oil are combusted on the surface of the catalyst, these inorganic fractions can remain on the catalyst, usually as oxides. Zinc oxide and phosphorus pentoxide have been found in prodigious amounts on catalyst surfaces. A zinc pyrophosphate glaze has been found to form on the catalyst surface after exposure to certain temperatures (80–82). This glaze completely masks large areas of the catalyst surfaces, preventing passage of the gases into the catalyst porous structure. Materials such as calcium inhibit the formation of this glaze.

Silicone residue introduced to gasoline with toluene plugged catalysts on vehicles (83). Also a manganese-based octane improver known as MMT has been shown to clog catalyst surfaces (84).

The accumulation of matter on the surface of the catalyst restricts gas passage into the catalyst by a mechanism known as pore mouth plugging (see Fig. 10b). It takes only a small amount of material on the surface of the catalyst to restrict the free passage of gases into and out of the active porous

catalyst layer.

Activated Layer Loss. Loss of the catalytic layer is the third method of deactivation. Attrition, erosion, or loss of adhesion and exfoliation of the active catalytic layer all result in loss of catalyst performance. The monolithic honeycomb catalyst is designed to be resistant to all of these mechanisms. There is some erosion of the inlet edge of the cells at the entrance to the monolithic honeycomb, but this loss is minor. The pelleted catalyst is more susceptible to attrition losses because the pellets in the catalytic bed rub against each other. Improvements in the design of the pelleted converter, the surface hardness of the pellets, and the depth of the active layer of the pellets also minimize loss of catalyst performance from attrition in that converter.

Oxygen Sensor and the Closed Loop Fuel Metering System

The first commercial application of the TWC catalyst and closed loop fuel metering system using an oxygen sensor came with the introduction of the 1977 Volvo for the California market (85–88). This catalyst was developed by Engelhard. Other car companies introduced the system the following year, and in the 1990s almost 100% of U.S. and Canada passenger cars and light-duty vehicles utilize it (89). The fully developed system is described in the literature (90) (see SENSORS).

The function of the oxygen sensor and the closed loop fuel metering system is to maintain the air and fuel mixture at the stoichiometric condition as it passes into the engine for combustion; ie, there should be no excess air or excess fuel. The main purpose is to permit the TWC catalyst to operate effectively to control HC, CO, and NO_x emissions. The oxygen sensor is located in the exhaust system ahead of the catalyst so that it is exposed to the exhaust of all cylinders (see Fig. 4). The sensor analyzes the combustion event after it happens. Therefore, the system is sometimes called a closed loop feedback system. There is an inherent time delay in such a system and thus the system is constantly correcting the air–fuel mixture cycles around the stoichiometric control point rather than maintaining a desired air–fuel mixture.

Oxygen Sensor. The oxygen sensor is also known as the lambda sonde or lambda sensor, from the greek letter used to denote the air–fuel ratio, and as an exhaust gas oxygen sensor (EGO). The lambda sensor responds in about 3 ms to a change in exhaust gas composition, yielding a voltage signal that is monitored by the computer control unit. The signal variation is from 50 mV at $\lambda = 1.05$ (lean) to 900 mV at $\lambda = 0.99$ (rich) as shown in Figure 11. The voltage changes sharply in the immediate vicinity of the stoichiometric ratio. A schematic of the lambda sensor is shown in Figure 12a. The sensor consists of a ceramic porous solid electrolyte that is ionically conductive at operating temperatures. The outside of the ceramic is coated with platinum electrodes: one electrode exposed to the exhaust gas, the second to outside air. Between these electrodes is a zirconia solid electrolyte. The voltage generated depends on the oxygen concentration difference between each electrode. The exhaust platinum electrode is a catalyst at the exhaust gas surface and equilibrates the mixture, consuming, by catalytic reaction, any unreacted oxygen and CO, HC, and H₂, yielding a net amount of oxygen present. Near the stoichiometric point, ie, $\lambda = 1.0$, the difference between atmospheric O₂ and exhaust O₂, and thus the voltage generated, approach a maximum. Thus the sensor signal is used to detect the stoichiometric point (91–95). In order for the sensor to function properly, it needs to be heated to above 350°C. To accommodate the need for quick catalyst heatup, a heated version of the oxygen sensor is used (96–99) (see Fig. 12b)

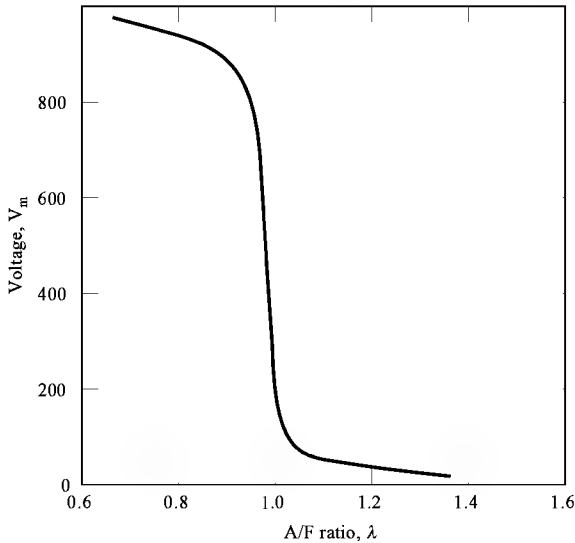


Fig. 11. The A–F ratio and the lambda sensor's voltage signal.

Courtesy of Robert Bosch.

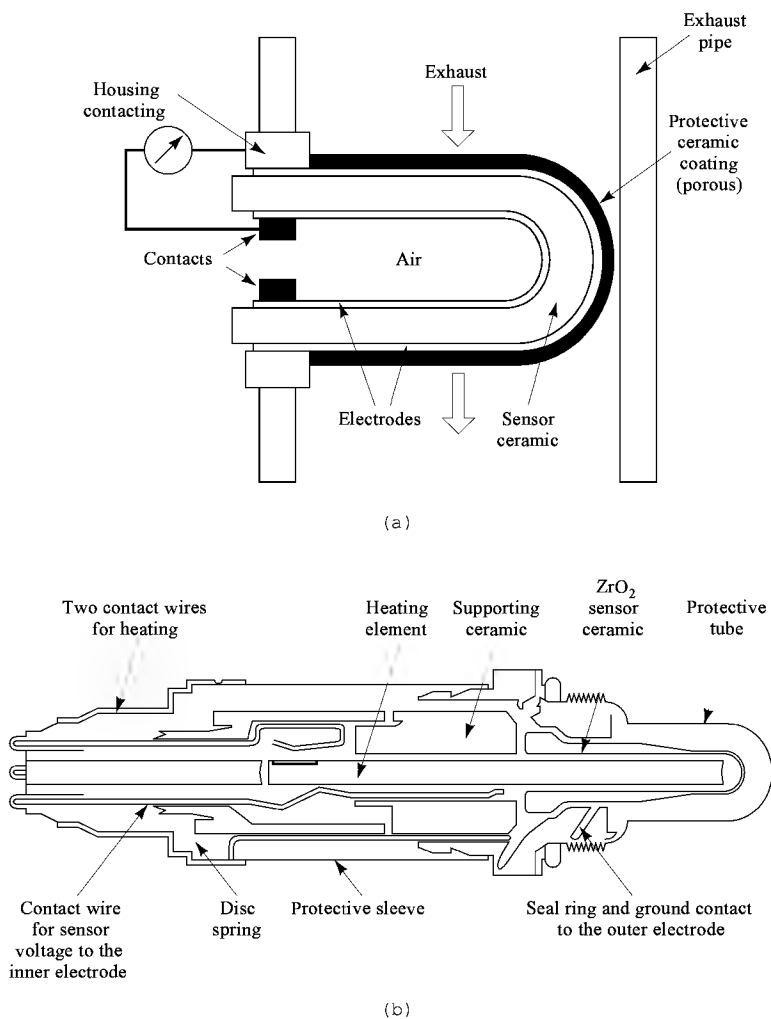


Fig. 12. Schematic of lambda sensor (a) in exhaust pipe and (b) internal components.

Courtesy of Robert Bosch.

Computer Controller. The computer controller takes inputs of speed, load, and temperature to assist the engine in cold starting and to select the optimal air fuel trim adjustment for optimal power, fuel economy, and emissions control. The control system uses the sensor signal as a switching device using the logic "rich! go lean!" or "lean! go rich!" Sophisticated adjustment logic is incorporated into the control unit to correct the air fuel mixture in anticipation of change.

The oxygen sensor closed loop system automatically compensates for changes in fuel content or air density. For instance, the stoichiometric air fuel mixture is maintained even when the vehicle climbs from sea level to high altitudes where the air density is lower.

The oxygen sensor is, however, fooled by certain gases. For instance, methanol combustion results in higher H₂ content in the exhaust, and because H₂ has a much higher rate of diffusion than other exhaust gases, a higher concentration is registered at the exhaust side electrode of the sensor than is actually present in the exhaust, causing a richer signal. The computer controller would then adjust the air fuel to a leaner setting. This factor has to be compensated for when using fuels such as methanol.

Closed loop control has been designed for both carburetors and fuel injection metering systems. The latter are used in almost all 1990 models. Two types of fuel metering exist: a single fuel injector to serve all cylinders, called single-point fuel injection; and fuel injectors for each cylinder, called multipoint fuel injection. The multipoint fuel injection systems may be continuous or individual electrically activated. An electronic fuel injection valve is located in the inlet manifold just ahead of each inlet valve. All valves are connected in parallel and open for a calibrated time as called for by the computer controller. The injected fuel is swept into the cylinder along with the air. The latest development is sequential fuel injection that meters fuel to each cylinder according to the firing order.

All closed loop control systems must measure the amount of air needed under all conditions of engine demand. Air measurement is most often done using a hot wire anemometer, usually referred to as a mass air meter (99,100).

The performance of the catalytic converter is affected by the conditions of air fuel control provided by the fuel metering system. A slowly responding fuel metering system can dramatically decrease the conversion efficiency of the converter compared to a fast response multipoint fuel injection system.

On-Board Diagnostics. State of California regulations require that vehicle engines and exhaust emission control systems be monitored by an on-board system to assure continued functional performance. The program is called OBD-II, and requires that engine misfire, the catalytic converter, and the evaporative emission control system be monitored (101). The U.S. EPA is expected to adopt a similar regulation.

One system for measuring catalyst failure is based on two oxygen sensors, one located in the normal control location, the other downstream of the catalyst (102,103). The second O₂ sensor indicates relative catalyst performance by measuring the ability to respond to a change in air fuel mixture. Other techniques using temperatures sensors have also been described (104–107). Whereas the dual O₂ sensor method is likely to be used initially, a criticism of the two O₂ sensors system has been reported (44) showing that properly functioning catalysts would be detected as a failure by the method.

Oxidation Catalyst. An oxidation catalyst requires air to oxidize unburned hydrocarbons and carbon monoxide. Air is provided with an engine driven air pump or with a pulse air device. Oxidation catalysts were used in 1975 through 1981 models but thereafter declined in popularity. Oxidation catalysts may be used in the future for lean burn engines and two-stroke engines.

The oxidation catalyst (OC) operates according to the same principles described for a TWC catalyst except that the catalyst only oxides HC, CO, and H₂. It does not reduce NO_x emissions because it operates in excess O₂ environments. One concern regarding oxidation catalysts was the ability to oxidize sulfur dioxide to sulfur trioxide, because the latter then reacts with water to form a sulfuric acid mist which is emitted from the tailpipe. The SO₂ emitted has the same ultimate fate in that SO₂ is oxidized in the atmosphere to SO₃ which then dissolves in water droplets as sulfuric acid.

Some OCs were of the monolithic honeycomb type, but all General Motors cars used pelleted OCs. For a period in the late 1970s and throughout the 1980s, both TWC and OC were used in a dual-bed catalyst. Oxygen needed for OC performance was provided by an engine driven air pump or a reed

valve (pulse air valve) positioned in the exhaust pipe or manifold. In the case of dual-bed catalysts, air was injected between the TWC and OC. OCs utilize Pt and/or Pd but not Rh as a rule. The OC is subjected to the same exhaust environments as the TWC, except the air/fuel ratio is different (108,109).

Exhaust Gas Recirculation. In one method of NO_x emission control, exhaust gas is fed back into the inlet manifold and mixed with the fuel and inlet air. The resultant mixture upon combustion in the cylinder results in lower peak combustion temperature and lower NO_x formation because the reaction of $\text{N}_2 + \text{O}_2 \rightarrow \text{NO}_x$ is strongly dependent on the combustion flame temperature (99,109–112). The degree of NO_x depression is dependent on the amount of exhaust gas recirculation (EGR) as shown in Figure 13. EGR provides a diluent gas having high molecular weight and CO_2 which absorbs heat. Also, EGR affects the flame speed of the mixture, and thus provides a certain antiknock quality to the combustion process. The impact of EGR on engine parameters has been detailed (113).

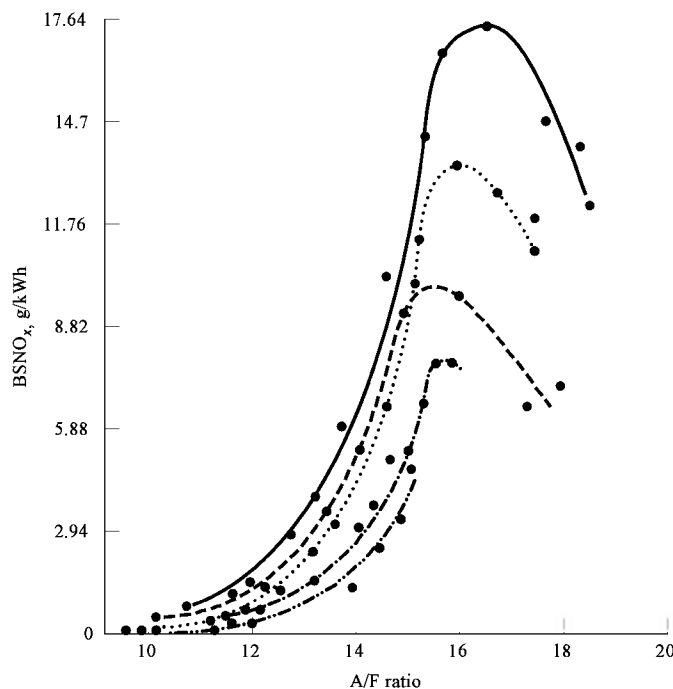


Fig. 13. Effect of EGR on brake specific NO_x (BSNO_x) production where (—) represents no EGR, (---) no EGR and a 20° retard, (.....), 5° EGR, (- · - · -) 10° EGR, and (- - - - -) 15° EGR.

EGR can seriously degrade engine performance, especially at idle, under load at low speed, and during cold start. Control of the amount of EGR during these phases can be accomplished by the same electronic computer controller used in the closed loop oxygen sensor TWC system. Thus the desired NO_x reduction is achieved while at the same time retaining good driveability.

Other Emissions Control

Evaporative Emission. Fumes emitted from stored fuel or fuel left in the fuel delivery system are also regulated by U.S. EPA standards. Gasoline consists of a variety of hydrocarbons ranging from high volatility butane (C-4) to lower volatility C-8 to C-10 hydrocarbons. The high volatility HCs are necessary for cold start, and are especially necessary for temperatures below which choking is needed to start the engine. Stored fuel and fuel left in the fuel system evaporates into the atmosphere.

Evaporative emissions sources are as follows. (1) Diurnal evaporative emissions are those that occur as ambient temperature fluctuates between daily high and low. The actual quantities of loss depend on the composition of the fuel and the daily temperatures. High concentrations of high volatility C-4 result in high vapor pressure and high evaporative emissions. (2) When the engine is turned off the engine and exhaust system heat up the fuel lines. This is known as hot soak, and fuel left in the fuel metering system or inlet manifold evaporates. (3) Evaporation also results from running loss, which occurs during operation at unusual conditions such as at low speeds on a hot day when the rate of vapor generation exceeds the rate that the engine can consume the vapor. (4) Evaporative loss source occurs during fueling as the liquid fuel displaces the fuel tank vapor.

Fuel vapor pressure, or Reid vapor pressure (RVP) is predominantly influenced by the amount of C-4 in the fuel. California, other of the United States, and the EPA have or are restricting fuel RVP for both seasonal and areas (latitude). These restrictions greatly reduce all gasoline evaporative losses.

The EPA regulation prohibits more than 2 grams per test to escape into the atmosphere (114). The test consists of a diurnal cycle of 1 hour where the temperature of the fuel is raised from 15.6 to 28.9°C during a 17 mile run on a chassis dynamometer. An immediate hot soak in a shed enclosure follows the dynamometer run.

The common method of controlling evaporative emission is an activated charcoal canister that connects to the intake manifold. When the engine is shut off, the valve permits hydrocarbon fumes to be absorbed and stored by the activated charcoal. When the engine is operating, the stored hydrocarbons are purged back into the manifold where they enter the combustion process. A new California regulation also controls running loss emissions of hydrocarbons to 0.03 g/km . There is discussion as to whether control of fueling emissions should be on-board the vehicle or incorporated into the fueling station system (115–118). As of 1993 the control is in the fueling station system.

Crankcase Emissions. Exhaust gases are also found in the crankcase. The principal source of these exhaust gases results from what is known as blow-by or gases from the combustion chamber, past the piston rings to the crankcase, or from the intake and exhaust valve mechanisms. As the engine wears, blow-by increases. Control systems are required to feed these gases back into the inlet manifold so that the hydrocarbons and carbon monoxide contained are consumed in the combustion process. The control device used is called a positive crankcase ventilation (PCV) valve. This device or one of similar function is required by law (119–121). Formerly, these gases were ventilated to the atmosphere through the engine breather tube. The test to measure crankcase emissions is defined in Reference 121.

Alternative Fuels

Under the National Energy Policy Act of 1992 nonpetroleum-based transportation fuels are to be introduced in the United States. Such fuels include natural gas (see *GAS, NATURAL*), liquefied petroleum gas (qv) (LPG), methanol (qv), ethanol (qv), and hydrogen (qv), although hydrogen fuels are not expected to be a factor until after the year 2000 (see also *ALCOHOL FUELS*; *HYDROGEN ENERGY*).

Natural Gas. Natural gas, an abundant fuel resource in the United States, has sufficient reserves to fuel over 10×10^6 U.S. vehicles per year for the next 50 years (122). Natural gas is used in two forms as a transportation fuel compressed or liquefied at low temperatures. Tanks for the storage of compressed natural gas are heavy and larger in volume than for liquid fuels. However, the added cost is offset by an expected lower pump price compared to gasoline (123). Whereas the lack of public natural gas fueling stations and other factors make natural gas more attractive for fleet vehicles in the United

States, small compressors for overnight natural gas refueling at household sites have been developed. Natural gas has a high antiknock quality, and engines designed for this fuel can have a higher compression ratio, yielding higher power and fuel economy. A disadvantage of compressed natural gas is that the driving range for a fuel load is usually 50% of that of gasoline. Work on an adsorptive system of natural gas storage at reduced storage pressures has been announced that should reduce compressed tank costs and may be able to improve range per fueling (124). Dual fuel systems have been developed in which the engine can run on either natural gas or gasoline.

Emission control for natural gas fueled engines consists of either the same principal components as used for gasoline or those designed for lean combustion. No evaporative emission control is required.

The catalysts for natural gas run engines are designed differently, however, because a primary portion of the exhaust from natural gas combustion is methane, and the catalysts are based on Pd rather than Pt. Spark ignited engines calibrated for natural gas usually are calibrated at stoichiometric conditions. Diesel engines re-engineered for natural gas are usually spark ignited and calibrated very lean. Pd–Rh is used for TWC catalysts for the stoichiometric engine; Pd oxidation catalysts are used for the lean re-engineered diesel engine. Both the U.S. EPA and the California standards exclude methane in the calculation of hydrocarbon emissions because methane has low reactivity in the atmosphere and has low ozone (qv) formation potential (125), and natural sources produce large quantities of methane. Methane, however, is a strong global warming gas and is regulated by a total hydrocarbon standard. The symmetrical methane molecule is the most refractory of the hydrocarbon family, and resists reactions on traditional auto exhaust catalysts at low temperatures. Pd catalysts need 450°C to oxidize methane but are nevertheless much better than all other precious metal catalysts. Methane is a poor reductant of NO_x (126,127). More recently, a Pd-based catalyst having poor dispersion (low metal surface area) was found to be superior to highly dispersed Pd for these reactions (128). Pd suffers temporary deactivation from SO₂ present in the exhaust. Natural gas contains very little sulfur. A small amount of mercaptan is added to give a characteristic odor to the fuel so that leaks in the distribution system can be detected. Pd-based catalysts therefore maintain performance much better than when used with gasoline (129–131).

LPG. LPG could be a principal alternative transportation fuel if its other uses were displaced by natural gas. A significant number of LPG fueling stations are located throughout the United States. LPG is a liquid fuel and does not suffer the same driving range problem as natural gas. Because LPG vapor pressure is high, the storage tank has to withstand 2800 kPa (400 psi).

The emission control system for LPG is the same as is used for gasoline fueled engines with the exception of the fuel metering system. No evaporative emission system is required. Both Pt–Rh and Pd–Rh catalysts are good for emission control of LPG fuel exhaust. Pt provides the lowest light off temperature for C₃H₈. The sulfur content of LPG is also very low so that Pd catalysts perform very well.

Methanol. Methanol is a liquid fuel made from natural gas, but can also be made from coal (qv), and the U.S. has huge coal resources. Engines designed for 100% methanol usually take advantage of the properties of methanol to provide high power. Methanol has high antiknock properties and can be used in high compression engines. Although methanol has a lower energy content, it has a higher latent heat of vaporization than gasoline. Evaporation in the inlet manifold cools and densifies the air–fuel charge to the engine, thus increasing the energy charge to the engine. Theoretically, therefore, the engine could be smaller and still provide the same power as a gasoline fueled engine. The disadvantage of 100% methanol engines is cold starting of the engine. Methanol has very low vapor pressure at cold temperatures, so there is insufficient vapor phase to provide a combustible mixture with a choked amount of air to start the engine.

Addition of 15% gasoline to methanol to produce M85 fuel is an alternative. At temperatures above –6.7°C, reliable ignition of M85 fuel occurs because the gasoline provides the vapor phase necessary for ignition under choked condition.

Engines are also designed to use either gasoline or methanol and any mixture thereof (132–136). Such a system utilizes the same fuel storage system, and is called a flexible fueled vehicle (FFV). The closed loop oxygen sensor and TWC catalyst system is perfect for the flexible fueled vehicle. Optimal emissions control requires a fuel sensor to detect the ratio of each fuel being metered at any time and to correct total fuel flow.

The principal hydrocarbon emissions from a methanol fueled car are methanol and formaldehyde. Formaldehyde is a carcinogen and regulated to very low levels by California and the U.S. EPA (0.015 g/m for light-duty vehicles (LDV)) and standards going to 0.008 g/m for California ultra low emission vehicle (ULEV) standards). Uncontrolled formaldehyde emissions from a methanol fueled light-duty vehicle are on the order of 0.60 g/m. Most TWC catalysts destroy formaldehyde effectively after operating temperature is reached, ie, about 1 to 2 minutes after the engine is started. However, during the 1 to 2 minutes that the catalyst is heating, about 0.05 to 0.1 g/m of formaldehyde is passed unreacted through the catalyst. The emission control system for either 100% methanol or for M85 must have catalysts located as close as possible to the manifold to heat up faster. Specially designed Pd–Rh TWC catalysts and close coupled manifold catalysts are being utilized for FFV to minimize formaldehyde emissions as well as to meet the other emission standards (136). An electrically heated catalyst system obtained very low formaldehyde emissions from a methanol fueled vehicle (137).

Future Engines and Emission Control Systems

Two engines are under development as of this writing: the two-stroke engine and the lean burn engine. The driving forces behind this development are fuel economy and global warming (see ATMOSPHERIC MODELS).

The National Highway and Safety Administration regulates vehicle fuel consumption through the Corporate Average Fuel Economy (CAFE). When vehicle manufacturers submit vehicles for certification, fuel consumption values must be given for a specified city and highway driving cycle (138). The average must be equal to or more than 11.6 km/L (27.5 mpg). Carbon dioxide is a principal global warming gas and conservation actions, such as increased CAFE, are thought to be a measure to stave off the buildup of CO₂ in the atmosphere.

Whereas automobile companies continue to improve the four-stroke engine, making it more efficient and more powerful to run vehicles that are smaller and lighter and have sufficient space, the two-stroke and the lean burn engines are also being studied.

Two-Stroke Engine. The two-stroke engine has the potential of delivering the same power as the four-stroke engine, but has fewer moving parts, lower weight, and a smaller engine volume, ie, it would be ideal for smaller, lighter, more fuel efficient vehicles having smaller engine compartments. Baseline emissions of two-stroke engines are very high in hydrocarbons. For example, a four-stroke engine may have 250 ppm hydrocarbons in its exhaust; the two-stroke engine could have 5000 ppm or greater. Designs that use inlet valves, fuel injection, and air compression have reduced baseline emissions, but have added complexity to the original simple two-stroke engine (139–147).

The operating air–fuel mixture of the two-stroke engine designs range from 1.3 to 2.0 stoichiometric. This lean mixture plus the characteristic internal exhaust gas recirculation lowers the peak combustion temperatures and results in low NO_x formation.

Emission control systems for two-stroke engines depend heavily on an efficient oxidation catalyst. These may be based on Pt and/or Pd. Higher lube oil consumption characteristics of two-stroke engines may result in modification to the lube oil or require the development of oxidation catalysts more resistant to lube oil ash compounds.

Lean Burn Engine. An engine calibrated at >18:1 A/F ratio, ie, very lean, encounters what is known as the lean limit (148). At this point, ignition of the air–fuel mixture becomes more difficult to achieve and subsequent propagation of the flame more difficult to sustain. Engineering solutions to both problems are diametrically opposed. At A/F ratios even more lean, combustion within the cylinder deteriorates and hydrocarbon emissions increase. Another result is loss of power. However, the production of NO_x is decreased at leaner A/F mixtures, and more importantly, thermal efficiency is improved and specific fuel consumption decreases. If engine designs could push back the lean limit, a practical lean burn engine may be a reality (149). To overcome the poor power and performance of a lean burn engine, a partial lean burn system was designed. The engine operated lean during idle and low cruise engine operating conditions. When acceleration was required, the engine adjusted to the stoichiometric air–fuel ratio. Toyota marketed a lean burn engine in Japan in the late 1980s and for a period in Europe without much customer acceptance (150). More recently, since 1990, several automobile manufacturers, such as Honda, Mitsubishi, Nissan, and Hyundai, have announced the development of the lean burn engine (151,152). All of these later developments used the Ford partial lean and partial stoichiometric calibration. Emission control for the lean burn–stoichiometric engine requires a TWC catalyst that can operate as a TWC catalyst at stoichiometric air–fuel mixtures and as an oxidation catalyst when the engine is operated lean.

Work on the development of the full time lean burn engine continues with efforts to push back the lean burn limit. A problem recognized by the developers is that although low basic engine NO_x emissions are possible, it is not yet possible to meet the NO_x emissions required by California and the

Tier II emission levels of the U.S. Clean Air Act of 1990 (see Table 1).

There has been a growing demand for a lean NO_x catalyst in order to decrease the relatively low NO_x emission of the lean burn engine sufficiently to meet the future standards. Lean NO_x catalysts have been developed based on zeolites (see MOLECULAR SIEVES). Cu-promoted ZSM-5 zeolite has shown ability to reduce NO_x in an exhaust having excess oxygen at an efficiency of 30 to 50% (153). Durability is not proven. Research has revealed that certain hydrocarbons are preferred for the reduction of NO_x , and that CO and H_2 apparently do not reduce NO_x over such lean NO_x catalysts (154). Considerable effort is being expended to develop a practical lean NO_x catalyst system (155–159).

Emission Control Technologies. The California low emission vehicle (LEV) standards has spawned investigations into new technologies and methods for further reducing automobile exhaust emissions. The target is to reduce emissions, especially HC emissions, which occur during the two minutes after a vehicle has been started (53). It is estimated that 70 to 80% of nonmethane HCs that escape conversion by the catalytic converter do so during this time before the catalyst is fully functional.

Technologies being investigated include improved catalysts that can be located closer to the exhaust manifold (high temperature resistant), reduced thermal mass in the exhaust system, use of alternative fuels, use of air pumps to assist catalyst light off, and engine adjustments that yield higher exhaust temperatures for this period. Three new technologies are undergoing intense development.

Electrically Heated Catalyst. An electrically heated catalyst (EHC) is a electrical resistive metallic foil or matrix which, like a toaster, will heat upon the application of voltage. Two basic types are being developed. One is based on a metal foil; the other is based on sintered metal powder. Figure 14a shows a cross-sectional view of a metal foil type, and Figure 14b shows a cross-sectional view of a the sintered metal powder type. A voltage, typically 12 V or 24 V supplied from the automobile battery, is applied across the electrodes for about 15 to 20 seconds, in order to heat the foil or metal matrix from ambient temperature to 300–600°C in this time period. The metal foil or metal matrix surfaces are covered with a catalytic layer as described for the monolithic converter. Thus, when the metal surface achieves 300°C, catalyst light-off occurs. The system usually employs an air pump so that there is sufficient oxygen present to consume significant quantities of CO and HCs, resulting in an additional temperature rise.

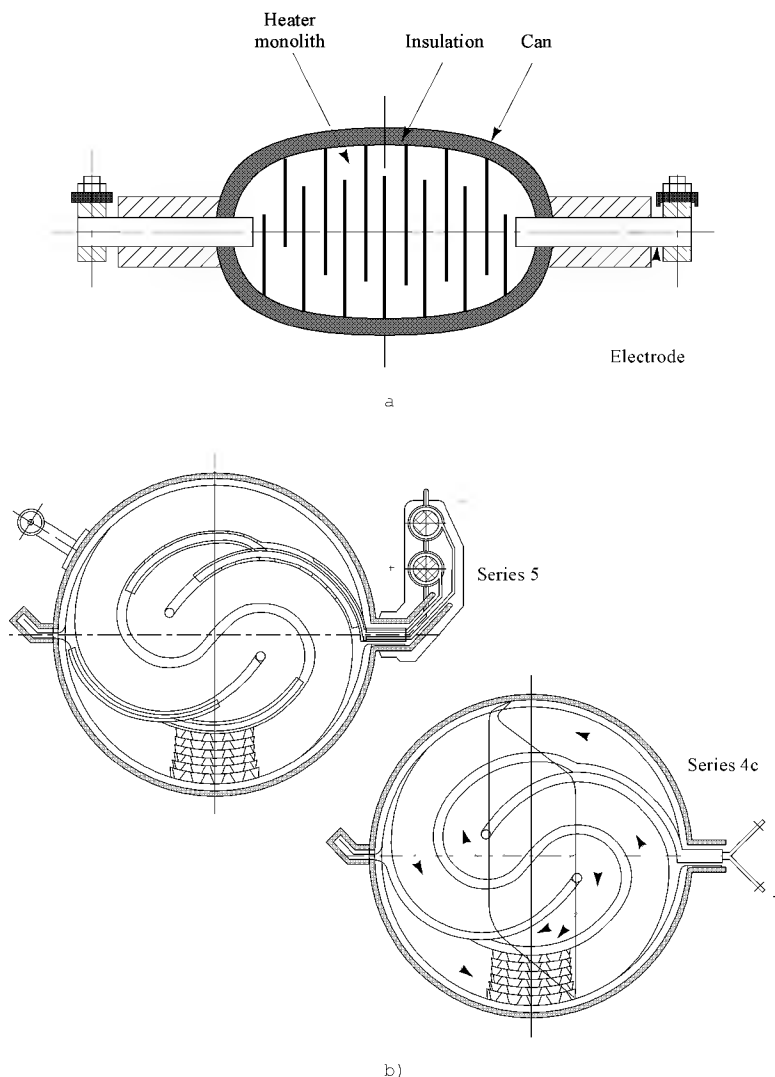


Fig. 14. Cross-sectional schematics of electrically heated catalyst (EHC) for emission control (a) extruded sintered metal powder EHC (160); (b) two sintered metal foil EHCs.

Courtesy of Emitec GmbH.

The main converter, which is located downstream of the EHC, heats to functional temperature much more quickly because of catalytic combustion of exhaust gases that would otherwise pass unconverted through the catalyst during the cold start period. The EHC theoretical power required for a reference case (161) was 1600 watts to heat an EHC to 400°C in 15 s in order to initiate the catalytic reactions and obtain the resultant exotherm of the chemical energy contained in the exhaust. Demonstrations have been made of energy requirements of 15–20 Wh and 2 to 3 kW of power (160,161). Such systems have achieved nonmethane HC emissions below the California ULEV standard of 0.025 g/km. The principal issues of the EHC are system durability, battery life, system complexity, and cost (137,162–168).

Hydrocarbon Trap System. The concept of a hydrocarbon trap or adsorber system is based on molecular sieve hydrocarbon adsorber systems. The temperatures at which hydrocarbon adsorption takes place exist in the auto engine exhaust system during the period of cold start of an automobile when the catalytic control system has not yet reached functional temperature. Zeolites have been reportedly useful for hydrocarbon adsorption (53,169). Zeolites desorb hydrocarbons at temperatures of 400°C, ie, once the catalytic control system is functional. Therefore, hydrocarbons adsorbed by the zeolite can also be desorbed then oxidized by a catalyst. Methods to accomplish cold start hydrocarbon adsorption, heatup of the main catalyst, and desorption have been identified. Some of these systems use exhaust pipe valves to divert the exhaust gases to the hydrocarbon trap for the low temperature portion, and by-pass the gases around the trap after the main catalyst has heated up. One device that uses a heat exchanger is shown in Figure 15 (44). The Si–Al ratio in the zeolite is important, and by lowering the alumina content, the zeolite is rendered more hydrophobic and more able to adsorb

hydrocarbons. Adsorption of 50 to 60% of the cold start hydrocarbons have been demonstrated (44,169–171). Zeolites do not efficiently adsorb C-2 to C-4 hydrocarbons.

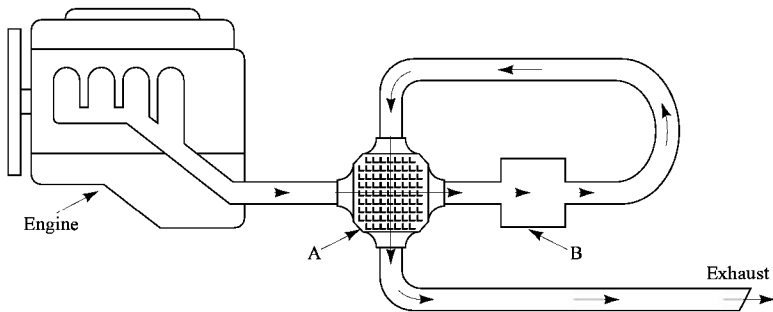


Fig. 15. Low hydrocarbon emission control system utilizing a cross-flow heat exchanger TWC catalyst, A, and a zeolite-based hydrocarbon absorber system. Cold start HCs are absorbed by the hydrocarbon trap, B, until the cross-flow heat exchanger catalyst is hot enough to destroy the HCs that subsequently are eluted from the HC trap.

Exhaust Gas Igniter. Exhaust gases during cold start have high concentrations of CO and HCs because of air restriction caused by choking. Using a severe choke strategy and addition of air by a pump the exhaust gas air mixture is flammable. The Ford Exhaust Gas Igniter incorporates a spark igniter in the exhaust line to ignite and sustain the flame. The flame heats up the main catalyst in 15 to 20 seconds, and in the process consumes those hydrocarbons and carbon monoxide that would otherwise pass through the cold catalyst (172,173). The system uses energy available in the exhaust gases during the initial seconds of engine starting. On the other hand, the choked condition is extreme, and there are associated engine concerns.

BIBLIOGRAPHY

"Automotive Exhaust Control" in *ECT* 2nd ed., Vol. 2, pp. 814–839, by W. B. Innes and K. Tsu, American Cyanamid Co.; "Exhaust Control, Automotive" in *ECT* 3rd ed., Vol. 9, pp. 494–510, by C. M. Heinen, Chrysler Corp.

1. M. P. Walsh, *Plat. Met. Rev.* **30**(3), 106–115 (1986).
2. C. Csere, *Car Driver*, 60–64 (Jan. 1988).
3. *Code of Federal Regulations*, Title 40, Part 86, Appendix IV, AMA Mileage Accumulation Route, Washington, D.C., July 1992, p. 951.
4. *1991 SAE Handbook*, Vol. 3. Society of Automotive Engineers, Warrendale, Pa., 1991, Chapt. 25.
5. A. L. Boehman, S. Niksa, and R. J. Moffatt, *A Comparison of Rate Laws for CO Oxidation Over Pt on Alumina*, SAE 930252, Society of Automotive Engineers, Warrendale, Pa., 1993.
6. J. T. Kummer, *Prog. Ener. Combust. Sci.* **6**, 177–199 (1981).
7. B. A. D'Alleva, *Procedure and Charts for Estimating Exhaust Gas Quantities and Composition*, General Motors Laboratory Report 372, Warren, Mich., 1960.
8. N. A. Heinen and D. J. Patterson, *Combustion Engine Economy, Emissions, and Controls*, University of Michigan, Ann Arbor, Mich., Chapt. 5, 1992.
9. L. Eltinge, *Fuel-Air Ratio and Distribution from Exhaust Gas Composition*, SAE 680114, Society of Automotive Engineers, Warrendale, Pa., 1968.
10. J. B. Edwards, *Combustion: The Formation and Emissions of Trace Species*, Ann Arbor Science Publishers, Ann Arbor, Mich., 1974, p. 16.
11. J. D. Benson and co-workers, *Effects of Gasoline Sulfur on Mass Exhaust Emissions—Auto/Oil Air Quality Improvement Research Program*, SAE 912323, Society of Automotive Engineers, Warrendale, Pa., 1991.
12. *Auto/Oil Air Quality Improvement Research Program*, SAE SP-920, Society of Automotive Engineers, Warrendale, Pa., 1992, 16 pp.
13. P. Blumberg and J. T. Kummer, *Combust. Sci. Technol.* **4**, 73 (1971).
14. N. Pelz and co-workers, *The Composition of Gasoline Engine Hydrocarbon Emissions—An Evaluation of Catalytic and Fuel Effects*, SAE 902074, Society of Automotive Engineers, Warrendale, Pa., 1990.
15. P. R. Shore, D. T. Humphries, and O. Haddad, *Speciated Hydrocarbon Emissions from Aromatic, Olefinic and Paraffinic Model Fuels*, SAE 930373, Society of Automotive Engineers, Warrendale, Pa., 1993.
16. G. J. den Otter, R. E. Malpas, and T. D. B. Morgan, *Effect of Gasoline Reformulation on Exhaust Emissions in Current European Vehicles*, SAE 930372, Society of Automotive Engineers, Warrendale, Pa., 1993.
17. J. Laurikko and N.-O. Nyland, *Regulated and Unregulated Emissions from Catalyst Vehicles at Low Ambient Temperatures*, SAE 930946, Society of Automotive Engineers, Warrendale, Pa., 1993. Good reference for low ambient temperature emissions.
18. P. R. Shore and R. S. de Vries, *On-Line Hydrocarbon Speciation Using FTIR and CIMS*, SAE 922246, Society of Automotive Engineers, Warrendale, Pa., 1992.
19. D. W. Wendland and W. R. Matthes, *Visualization of Automotive Catalytic Converter Internal Flows*, SAE 861554, 1986; D. W. Wendland, P. L. Sorrell, and J. E. Kreucher, *Sources of Monolithic Catalytic Converter Pressure Loss*, SAE 912372, Society of Automotive Engineers, Warrendale, Pa., 1991.
20. H. Weltans, H. Bressler, and P. Krause, *Influence of Catalytic Converters on Acoustics of Exhaust Systems for European Cars*, SAE 910836, Society of Automotive Engineers, Warrendale, Pa., 1991.
21. D. W. Wendland, W. R. Matthes, and P. L. Sorrell, *Effect of Header Truncation on Monolith Converter Emission Control Performance*, SAE 922340, Society of Automotive Engineers, Warrendale, Pa., 1992.
22. H. Weltans and co-workers, *Optimization of Catalytic Converter Gas Flow Distribution by CFD Prediction*, SAE 930780, Society of Automotive Engineers, Warrendale, Pa., 1993.
23. I. Sword and co-workers, SAE 930943, Society of Automotive Engineers, Warrendale, Pa., 1993.
24. J. Howitt, *Thin Wall Ceramics as Monolithic Catalyst Supports*, SAE 910611, Society of Automotive Engineers, Warrendale, Pa., 1991.
25. L. S. Socha, J. P. Day, and E. M. Barnett, *Impact of Catalytic Support Design Parameters on FTP Emissions*, SAE 892041, Society of Automotive Engineers, Warrendale, Pa., 1989.
26. H. Yamamoto and co-workers, *Warm-Up Characteristics of Thin Wall Honeycomb Catalysts*, SAE 910611, Society of Automotive Engineers, Warrendale, Pa., 1991.
27. J. P. Day and L. S. Socha, Jr., *The Design of Automotive Catalyst Supports for Improved Pressure Drop and Conversion Efficiency*, SAE 910371, Society of Automotive Engineers, Warrendale, Pa., 1991.
28. M. Luoma, P. Lappi, and R. Lylykangas, *Evaluation of High Cell Density Z-Flow Catalyst*, SAE 930940, Society of Automotive Engineers, Warrendale, Pa., 1993. Good reference for mass-transfer limited model reactions.
29. K. Otto, W. B. Williamson, and H. Gandhi, *Ceram. Eng. Sci. Proc.* **2**(6), (May–June, 1981). Good review of various catalyst deactivation processes.
30. R. J. Farrauto and M. C. Hodson, "Catalyst Characterization," *Encyclopedia of Physical Science and Technology*, Vol. 2, Academic Press, Inc., New York, 1987.
31. J. T. Kummer, "The Use of Noble Metals in Automobile Exhaust Catalyst," *Chicago Symposium*, 1985.

32. J. E. Kubsh, J. S. Rieck, and N. D. Spencer, in A. Crucq, Ed., *Catalysis and Automotive Pollution Control II*, Elsevier, New York, 1991, p. 125.
33. J. K. Hockmuth and co-workers, *Hydrocarbon Traps for Controlling Cold Start Emissions*, SAE 930739, Society of Automotive Engineers, Warrendale, Pa., 1993.
34. Y. T. Yu Yao, *Indust. Eng. Chem. Prod. Res. Dev.* **19**, 293 (1980).
35. H. C. Yao, H. K. Stephen, and H. S. Gandhi, *J. Catal.* **61**, 547 (1980).
36. H. C. Yao, S. Japar, and M. Shelef, *J. Catal.* **50**, 407 (1977).
37. S. H. Oh, P. J. Mitchel, and R. M. Siewert, *J. Catal.* **132**, 287 (1991).
38. R. F. Hicks, C. Rigano, and B. Pang, *Catal. Lett.* **6**, 271 (1990).
39. T. Yamada, K. Kayano, and M. Funibike, *The Effectiveness of Pd for Converting Hydrocarbons in TWC Catalysts*, SAE 930253, Society of Automotive Engineers, Warrendale, Pa., 1993.
40. J. C. Dettling and W. K. Lui, *A Non-Rhodium Three-Way Catalyst for Automotive Applications*, SAE 920094, Society of Automotive Engineers, Warrendale, Pa., 1992.
41. C. N. Montreuil, S. D. Williams, and A. A. Adanczyk, *Modeling Current Generation Catalytic Converters: Laboratory Experiments and Kinetic Parameter Optimization—Steady State Kinetics*, SAE 920096, Society of Automotive Engineers, Warrendale, Pa., 1992.
42. M. A. Harkonen and P. Talvitie, *Optimization of Metallic TWC Behavior and Precious Metal Costs*, SAE 920395, Society of Automotive Engineers, Warrendale, Pa., 1992.
43. J. T. Kummer, *J. Phys. Chem.* **90**, 4747 (1986).
44. G. B. Fisher and co-workers, *The Role of Ceria in Automotive Exhaust Catalysis and OBD-II Catalyst Monitoring*, SAE 931034, Society of Automotive Engineers, Warrendale, Pa., 1993.
45. S. Pelters, F. W. Kaiser, and W. Maus, *The Development and Application of Metal Supported Catalyst for Porsche's 911 Carrera 4*, SAE 890488, Society of Automotive Engineers, Warrendale, Pa., 1989.
46. F. W. Kaiser and S. Pelters, *Comparison of Metal Supported Catalysts with Different Cell Geometries*, SAE 910837, Society of Automotive Engineers, Warrendale, Pa., 1991.
47. P. Oser, *Novel AutoCatalyst Concepts and Strategies for the Future with Emphasis on Metal Supports*, SAE 880319, Society of Automotive Engineers, Warrendale, Pa., 1988.
48. K. Nishizawa and co-workers, *Metal Supported Automotive Catalysts for Use in Europe*, SAE 880188, Society of Automotive Engineers, Warrendale, Pa., 1988.
49. S. H. Oh and D. L. Van Ostrom, *A Three Dimensional Model for the Analysis of Transient Thermal and Conversional Characteristics of Monolithic Catalytic Converters*, SAE 880282, Society of Automotive Engineers, Warrendale, Pa., 1988.
50. J. C. W. Kuo, C. R. Morgan, and H. G. Lasson, *Mathematical Modelling of CO and HC Catalytic Converter Systems*, SAE 710289, Society of Automotive Engineers, Warrendale, Pa., 1971.
51. C. N. Montreuil, S. D. Williams, and A. A. Adanczyk, *Modeling Current Generation Catalytic Converters; Laboratory Experiments and Kinetic Parameter Optimization—Steady State Kinetics*, SAE 920096, Society of Automotive Engineers, Warrendale, Pa., 1992.
52. K. S. Creamer and J. H. Sanders, *Evaluation of a Catalytic Converter for a 3.73 kW Natural Gas Engine*, SAE 930221, Society of Automotive Engineers, Warrendale, Pa., 1993.
53. S. Tauster, Engelhard Corp., Iselin, N. J., 1992.
54. C. M. Friend, *Sci. Am.* **268**(4), 74–79 (Apr. 1993).
55. M. Golze and co-workers, *Phys. Rev. Lett.* **53**(8), 850–853 (1993).
56. R. J. Maddix, *Science* **V233**, 1159–1166 (Sept. 12, 1986).
57. G. Fisher and co-workers, "Mechanism of the Nitric Oxide–Carbon Monoxide–Oxygen Reaction Over a Single Crystal Rhodium Catalyst," in M. J. Philips and M. Ternan, eds., *Proceedings of the 9th International Congress on Catalysis*, Vol 3, *Characterization and Metal Catalysts*, Chemical Institute of Canada, Ottawa, 1988.
58. J. T. Yates, Jr., *Chem. Eng. News* **70**(13), 22–35 (Mar. 30, 1992).
59. J. C. Summers and S. A. Ansen, *J. Catal.* **58**, 131 (1979).
60. H. C. Yao and Y. F. Yu Yao, *J. Catal.* **86**, 254 (1984).
61. K. Ihara, H. Marakami, and K. Ohkubo, *Improvement of Three-Way Catalyst Performance by Optimizing Ceria Impregnation*, SAE 902168, Society of Automotive Engineers, Warrendale, Pa., 1990.
62. A. F. Duvell, R. R. Rojaram, H. A. Shaw, and T. J. Truex, in A. Crucq, ed., in Ref. 32, p. 139.
63. R. Hicks and co-workers, *J. Catal.* **122**, 296–306 (1990).
64. H. S. Gandhi, W. B. Williamson, and J. L. Bomback, *Appl. Catal.* **3**, 79–88 (1982).
65. E. Jobson and co-workers, *Deterioration of Three-Way Automotive Catalysts, Part I—Steady State and Transient Emission of Aged Catalyst*, SAE 930937, Society of Automotive Engineers, Warrendale, Pa., 1993.
66. S. Lundgren and E. Jobson, *Deterioration of Three-Way Automotive Catalysts, Part II—Oxygen Storage Capacity at Exhaust Conditions*, SAE 930944, Society of Automotive Engineers, Warrendale, Pa., 1993.
67. L. A. Carol, N. E. Newman, and G. S. Mann, *High Temperature Deactivation of Three-Way Catalyst*, SAE 892040, Society of Automotive Engineers, Warrendale, Pa., 1989.
68. N. Miyoshi and co-workers, *Development of Thermal Resistant Three-Way Catalyst*, SAE 891970, Society of Automotive Engineers, Warrendale, Pa., 1989.
69. C. D. Tyree, *Emission Level in Catalyst Temperature as a Function of Ignition-Induced Misfire*, SAE 920298, Society of Automotive Engineers, Warrendale, Pa., 1992.
70. N. Miyoshi and co-workers, *Development of Thermal Resistant Three-Way Catalysts*, SAE 891970, Society of Automotive Engineers, Warrendale, Pa., 1989.
71. M. P. Walsh, *The Advantages of Removing Lead from Gasoline and Using Catalytic Converters to Control Vehicle Exhaust Pollution*, sponsored by Corning Glass Works, 1983.
72. Needleman and co-workers, *New Engl. J. Med.* **300**(13), (Mar. 29, 1979).
73. A. M. Hochhauser and co-workers, *Effects of Gasoline Sulfur Level on Mass Exhaust Emissions—Auto/Oil Air Quality Improvement Research Program*, SAE 912323, Society of Automotive Engineers, Warrendale, Pa., 1991.
74. J. C. Summers and co-workers, SAE 920558, Society of Automotive Engineers, Warrendale, Pa., 1992.
75. **U.S. Pat. 4,552,733** (Nov. 12, 1985), C. Thompson, J. Mooney, C. D. Keith, and W. Mannion (to Engelhard Corp.)
76. I. Gottberg, E. Hogberg, and K. Weber, *Sulfur Storage and Hydrogen Sulphide Release From a Three-Way Catalyst Equipped Car*, SAE 890491, Society of Automotive Engineers, Warrendale, Pa., 1989.
77. J. C. Dettling and co-workers, *Control of H₂S Emissions from High-Tech TWC Converters*, SAE 900506, Society of Automotive Engineers, Warrendale, Pa., 1990.
78. S. E. Golunski and S. A. Roth, *Catal. Today* **9**, 109 (1991).
79. **U.S. Pat. 5196390** (Mar. 23, 1993), S. Tauster, J. Dettling, and J. Mooney (to Engelhard Corp.).
80. W. B. Williamson and co-workers, *Catalyst Deactivation Due to Glaze Formation from Oil-Derived Phosphorus and Zinc*, SAE 841406, Society of Automotive Engineers, Warrendale, Pa., 1984.

81. K. Inoue, T. Kurahashi, and T. Negishi, SAE 92064, Society of Automotive Engineers, Warrendale, Pa., 1992.
82. W. B. Williamson and co-workers, *Appl. Catal.* **15**, 277–292 (1985).
83. H. S. Gandhi and co-workers, "Affinity of Lead for Mobile Metals on Different Supports," *Surface Interface Analy.* **6**(4), (1984).
84. W. B. Williamson, H. S. Gandhi, and E. E. Weaver, *Effects of Fuel Additive MMT on Contaminant Retention and Catalyst Performance*, SAE 821193, Society of Automotive Engineers, Warrendale, Pa., 1982.
85. G. T. Engh and S. Wallman, *Development of the Volvo Lambda-Sond System*, SAE 770295, Society of Automotive Engineers, Warrendale, Pa., 1977.
86. J. J. Mooney, C. E. Thompson, and J. C. Dettling, *Three-Way Conversion Catalysts—Part of the New Emission Control System*, SAE 770365, Society of Automotive Engineers, Warrendale, Pa., 1977.
87. E. Koberstein, *Characterization of Multifunctional Catalysts for Automotive Exhaust Purifications*, SAE 770366, Society of Automotive Engineers, Warrendale, Pa., 1977.
88. P. Oser, *Catalyst Systems with and Emphasis on Three-Way Conversion and Novel Concepts*, SAE 790306, Society of Automotive Engineers, Warrendale, Pa., 1979.
89. K. Oblander, J. Abthoff, and H. D. Schuster. *Der Driewegkatalysator-eine Abgasreinigungstechnologie für Kraftfahrzeuge mit Ottomotoren (Three-Way Catalyst—An Exhaust Cleaning Technology for Automobiles with Internal Combustion Engines)*, VDI-Berichte 630, VDI-Verlag, Dusseldorf, Germany, 1984.
90. J. Abthoff and co-workers, *MTZ Motortechnische Zeitschrift* **53**, 11 (1992).
91. E. Logothetis, *Sci. Technol. Zirconia*, **3**, 388 (1981); R. E. Hettrich, W. A. Fate, and W. C. Vassell, *Oxygen Sensing by Electrochemical Pumping*, SAE 810433, Society of Automotive Engineers, Warrendale, Pa., 1981.
92. W. J. Fleming, *Zirconia Oxygen Sensor—An Equivalent Circuit Model*, SAE 800020, Society of Automotive Engineers, Warrendale, Pa., 1980.
93. E. Hamann, H. Manger, and L. Stencke, *Lambda Sensor with Y2O3—Stabilized ZrO2—Ceramic for Application in Automotive Emission Control Systems*, SAE 770401, Society of Automotive Engineers, Warrendale, Pa., 1977.
94. J. W. Butler and co-workers, *Fast Response Zirconia Sensor-Based Instrument for Measurement of the Air/Fuel Ratio of Combustion Exhaust*, SAE 840061, Society of Automotive Engineers, Warrendale, Pa., 1984.
95. I. Gorille, N. Rittmansberger, and P. Werner, *Bosch Electronic Fuel Injection with Closed Loop Control*, SAE 750368, Society of Automotive Engineers, Warrendale, Pa., 1975.
96. E. Hendricks, T. Vesterholm, and S. C. Sorenson, *Non Linear Closed Loop SI Engine Control Observers*, SAE 920237, Society of Automotive Engineers, Warrendale, Pa., 1992.
97. A. J. Beumont, A. D. Noble, and A. Scariobrick, "Adaptive Transient Air Fuel Ratio Control to Minimize Gasoline Engine Emissions," *Fisita Congress*, London, 1992.
98. M. J. Anderson, *A Feedback A/F Control System for Low Emission Vehicles*, SAE 930388, Society of Automotive Engineers, Warrendale, Pa., 1993.
99. *Automotive Handbook*, 2nd ed., Robert Bosch GmbH, Stuttgart, Germany, 1986, p. 442.
100. C. O. Probst, *Bosch Fuel Injection & Engine Management*, Robert Bentley, Cambridge, Mass., 1989.
101. Technical Support Document *Revisions to the Malfunction and Diagnostic System Requirements Applicable to 1994 and later New California Passenger Cars, Light-Duty Trucks, and Medium Duty Vehicles with Feedback Fuel Control Systems (OBD-II)*, California Air Resources Board, Sacramento, Sept. 14, 1989.
102. J. W. Koupal, M. A. Sabourin, and W. V. Clemmens, *Detection of Catalyst Failure On-Vehicle Using the Dual Oxygen Sensor Method*, SAE 91061, Society of Automotive Engineers, Warrendale, Pa., 1991.
103. W. Clemmens, M. Sabourin, and T. Rao, *Detection of Catalyst Failure Using On-Board Diagnostics*, SAE 900062, Society of Automotive Engineers, Warrendale, Pa., 1992.
104. S. H. Oh, *Thermal Response of a Monolith Catalyst Converter During Sustained Misfiring: A Computational Study*, SAE 881591, Society of Automotive Engineers, Warrendale, Pa., 1988.
105. W. Cai, Novel Sensors for On Vehicle Measurement of Emissions, Ph.D. dissertation University of Cambridge, Mass., 1992.
106. W. Cai and N. Collings, *A Catalytic Oxidation Sensor for the On-Board Detection of Misfire and Catalyst Efficiency*, SAE 922248, Society of Automotive Engineers, Warrendale, Pa., 1992.
107. N. Collings and co-workers, *A Linear Catalyst Temperature Sensor for Exhaust Gas Ignition (EGI) and On Board Diagnostics of Misfire and Catalyst Efficiency*, SAE 930938, Society of Automotive Engineers, Warrendale, Pa., 1993.
108. D. D. Bech and co-workers, *The Performance of Pd, Pt and Pd-Pt Catalysts in Lean Exhaust*, SAE 930084, Society of Automotive Engineers, Warrendale, Pa., 1993.
109. H. E. Jaaskelainen and J. S. Wallace, *Performance and Emissions of a Natural Gas-Fueled 16-Valve DOHC Four-Cylinder Engine*, SAE 930380, Society of Automotive Engineers, Warrendale, Pa., 1993.
110. A. A. Quader, *Why Intake Charge Dilution Decreases Nitric Oxide Emissions from Spark Ignited Engines*, SAE 710009, Society of Automotive Engineers, Warrendale, Pa., 1971.
111. J. A. Harrington and R. C. Shishu, *Zirconia Oxygen Sensor—An Equivalent Circuit Model*, SAE 730476, Society of Automotive Engineers, Warrendale, Pa., 1973.
112. E. A. Mayer, *Electro-Pneumatic Control Valve for EGR/ATC Actuation*, SAE 810464, Society of Automotive Engineers, Warrendale, Pa., 1981.
113. J. J. Gumbleton and co-workers, *Optimizing Engine Parameters with Exhaust Gas Recirculation*, SAE 740104, Society of Automotive Engineers, Warrendale, Pa., 1974.
114. Ref. 4, Chapt. 25.89, SAE J171.
115. H. M. Haskey, W. R. Cadman, and T. F. Liberty, *The Development of a Real-Time Evaporative Emission Test*, SAE 901110, Society of Automotive Engineers, Warrendale, Pa., 1990.
116. L. L. Lave, W. E. Wecher, W. S. Reis, and D. A. Ross, *Environ. Sci. Technol.* **24**, 8 (1990).
117. *The Effects of Temperature and Fuel Volatility on Vehicle Evaporative Emissions*, Report No 90-1, Concawe, Brussels, Belgium, 1990.
118. T. Cam, K. Cullen, and S. L. Baldus, *Running Loss Temperature Profile*, SAE 930078, Society of Automotive Engineers, Warrendale, Pa., 1993; H. M. Haskew, W. R. Cadman, and T. F. Liberty, *Evaporative Emissions Under Real-Time Conditions*, SAE 891121, Society of Automotive Engineers, Warrendale, Pa., 1989.
119. C. B. Tracy and W. W. Frank, *Fuels, Lubricants, and Positive Crankcase Ventilation System*, SAE PT112, 451, Society of Automotive Engineers, Warrendale, Pa., 1963–1966.
120. R. J. Templin, *Discussion of Reference 2*, SAE TP-6, 249, Society of Automotive Engineers, Warrendale, Pa., 1964.
121. Ref. 4, Chapt. 25.68, SAE J900.
122. A. Unich, R. M. Bada, and K. W. Lyons, *Natural Gas: A Promising Fuel for I.C. Engines*, SAE 930929, Society of Automotive Engineers, Warrendale, Pa., 1993. An extensive reference source.
123. American Gas Association (AGA), "Projected Natural Gas Demand From Vehicles Under the Mobile Sources Provisions of the Clean Air Act Amendments," *AGA Energy Analysis EA 1991–1991*, Chicago, 1991.
124. M. Samsa, *Potential for Compressed Natural Gas Vehicles in Centrally-Fueled Automobile, Truck and Bus Fleet Application*, Gas Research Institute, Chicago, 1991, pp. 44–61.
125. Technical staff report on Reactivity Adjustment Factors (RAF), California Air Resources Board, 1993.
126. J. J. White, *Low Emission Catalysts for Natural Gas Engines*, GRI Report 91-0214, Gas Research Institute, Chicago, Southwest Research Institute, San Antonio, SwRI 3178–22, 1991.
127. J. Klimstra, *Catalytic Converters for Natural Gas Engines—A Measurement and Control Problem*, SAE 872165, Society of Automotive Engineers,

Warrendale, Pa., 1987.

128. R. Hicks and co-workers, *Structure Sensitivity of Methane Oxidation over Platinum and Palladium*; *J. Catal.*, 280–306 (1990).

129. W. M. Burkmyre, W. E. Liss, and M. Church, *Natural Gas Converter Performance and Durability*, SAE 930222, Society of Automotive Engineers, Warrendale, Pa., 1993.

130. S. Subramanian, R. J. Kudla, and M. S. Chattha, *Treatment of Natural Gas Vehicle Exhaust*, SAE 930223, Society of Automotive Engineers, Warrendale, Pa., 1993.

131. J. E. Sinor and B. K. Bailey, *Current and Potential Future Performance of Ethanol Fuels*, SAE 930376, Society of Automotive Engineers, Warrendale, Pa., 1993.

132. J. J. Mooney, J. G. Hansel, and K. R. Burns, *Three-Way Conversion Catalysts on Vehicles Fueled with Ethanol–Gasoline Mixture*, SAE 790428, Society of Automotive Engineers, Warrendale, Pa., 1979.

133. R. K. Pfeffy, University of Santa Clara, personal communication, 1979.

134. C. L. Myring and co-workers, *Research and Development of Hyundai Flexible Fuel Vehicles (FFVs)*, SAE 930330, Society of Automotive Engineers, Warrendale, Pa., 1993.

135. T. Suga and Y. Hamazaki, *Development of Honda Flexible Fuel Vehicle*, SAE 922276, Society of Automotive Engineers, Warrendale, Pa., 1992.

136. J. K. Hochmuth and J. J. Mooney, *Catalytic Control of Emissions from M-85 Fueled Vehicles*, SAE 930219, Society of Automotive Engineers, Warrendale, Pa., 1993.

137. K. H. Hellman, G. K. Piotrowski, and R. M. Schaefer, *Start Catalyst Systems Employing Heated Catalyst Technology for Control of Emissions from Methanol-Fueled Vehicles*, SAE 93082, Society of Automotive Engineers, Warrendale, Pa., 1993.

138. Ref. 4, Chapt. 24.373.

139. R. Douglas and G. P. Blair, *Fuel Injection of a Two-Stroke Cycle Spark Ignition Engine*, SAE 820952, Society of Automotive Engineers, Warrendale, Pa., 1982.

140. D. Plohberger and co-workers, *Development of a Fuel-Injected Two-Stroke Gasoline Engine*, SAE 880170, Society of Automotive Engineers, Warrendale, Pa., 1988.

141. M. Nuti, *A Variable Timing Electronically Controlled High Pressure Injection System for 2S S.I. Engines*, SAE 900799, Society of Automotive Engineers, Warrendale, Pa., 1990.

142. G. E. Hundleby, *Development of a Poppet-Valved Two-Stroke Engine—The Flagship Concept*, SAE 900802, Society of Automotive Engineers, Warrendale, Pa., 1990.

143. G. P. Blair and co-workers, "The Reduction of Emissions and Fuel Consumption by Direct Air-Assisted Fuel Injection into a Two-Stroke Engine," *The 4th Graz Two-Wheeler Symposium*, Technical University, Graz, Austria, 1991.

144. J. Stokes and co-workers, *Development Experience of a Popped-Valved Two-Stroke Flagship Engine*, SAE 920778, Society of Automotive Engineers, Warrendale, Pa., 1992.

145. H. H. Huang and co-workers, *Improvement of Irregular Combustion of Two-Stroke Engine by Skip Injection Control*, SAE 922310, Society of Automotive Engineers, Warrendale, Pa., 1992.

146. C. Csere, *Car Driver*, 87–95, (June 1991).

147. *Auto. Eng.* **99**(7), 11–14 (July 1991).

148. F. Markus, *Car Driver*, 72–75 (Feb. 1992).

149. K. Oblander, J. Abthoff, and L. Fricker, "From Engine Testbench to Vehicle—An Approach to Lean Burn by Dual Ignition," *I. Mech E. C80/79*, (1979).

150. S. Matsushita and co-workers, *Development of the Toyota Lean Combustion System*, SAE 850044, Society of Automotive Engineers, Warrendale, Pa., 1985.

151. I-Y Ohm and co-workers, *Development of HMC Axially Stratified Lean Combustion Engine*, SAE 930879, Society of Automotive Engineers, Warrendale, Pa., 1993.

152. T. Inoue and co-workers, *Toyota Lean Combustion System—the Third Generation System*, SAE 930873, Society of Automotive Engineers, Warrendale, Pa., 1993.

153. M. Iwamoto and H. Hamada, *Catal. Today* **10**, 57 (1991).

154. B. H. Engler and co-workers, *Catalytic Reduction of NO_x with Hydrocarbons under Lean Diesel Exhaust Gas Conditions*, SAE 930735, Society of Automotive Engineers, Warrendale, Pa., 1993.

155. W. Held and co-workers, *Catalytic NO_x Reduction in Net Oxidizing Exhaust Gas*, SAE 900496, Society of Automotive Engineers, Warrendale, Pa., 1990.

156. T. J. Truex, R. A. Searles, and D. C. Sun. *Plat. Met. Rev.* **36**(1), 2–11 (Jan. 1992). A good NO_x reduction background reference.

157. G. Muramatsu and co-workers, *Catalytic Reduction of NO_x in Diesel Exhaust*, SAE 930135, Society of Automotive Engineers, Warrendale, Pa., 1993.

158. M. J. Heimrich and M. L. DeViney, *Lean NO_x Catalyst Evaluation and Characterization*, SAE 930736, Society of Automotive Engineers, Warrendale, Pa., 1993.

159. D. R. Monroe and co-workers, *Evaluation of a Cu-Zeolite Catalyst to Remove NO_x from Lean Exhaust*, SAE 930737, Society of Automotive Engineers, Warrendale, Pa., 1993.

160. L. S. Socha, D. F. Thompson, and P. S. Weber. *Reduced Energy and Power Consumption for Electrically Heated Extruded Metal Converters*, SAE 930383, Society of Automotive Engineers, Warrendale, Pa., 1993.

161. F. W. Kaiser and co-workers, *Optimization of an Electrically-Heated Catalytic Converter System—Calculations and Application*, SAE 930384, Society of Automotive Engineers, Warrendale, Pa., 1993.

162. M. J. Heimrich, S. Albu, and J. Osborne, *Electrically Heated System Conversion on Two Current Technology Vehicles*, SAE 910612, Society of Automotive Engineers, Warrendale, Pa., 1991.

163. I. Gottberg and co-workers, *New Potential Exhaust Gas Aftertreatment Technologies for Clean Car Legislation*, SAE 910849, Society of Automotive Engineers, Warrendale, Pa., 1991.

164. K. H. Hellman, C. K. Piotroski, and R. M. Shaefer, *Evaluation of Different Resistively Heated Technologies*, SAE 912382, Society of Automotive Engineers, Warrendale, Pa., 1991.

165. R. J. Hurley and co-workers, *Experiences with Electrically Heated Catalysts*, SAE 912384, Society of Automotive Engineers, Warrendale, Pa., 1991.

166. J. E. Kubsh and P. W. Lissuik, *Vehicle Emission Performance with an Electrically Heated Converter System*, SAE 912385, Society of Automotive Engineers, Warrendale, Pa., 1991.

167. L. S. Socha and D. F. Thompson, *Electrically Heated Extruded Metal Converter for Low Emission Vehicles*, SAE 920093, Society of Automotive Engineers, Warrendale, Pa., 1992.

168. O. Haddad and co-workers, *The Achievement of ULEV Emission Standards for Large High Performance Vehicles*, SAE 930389, Society of Automotive Engineers, Warrendale, Pa., 1993.

169. EP 424966 (1990), T. Minami and T. Nagase.

170. B. H. Engler and co-workers, *Reduction of Exhaust Gas Emission by Using Hydrocarbon Adsorber Systems*, SAE 930738, Society of Automotive Engineers, Warrendale, Pa., 1993.

171. M. J. Heimrich, L. R. Smith, and J. Kitowski, *Cold-Start Hydrocarbon Collection for Advanced Exhaust Emission Control*, SAE 920847, Society of

Automotive Engineers, Warrendale, Pa., 1992.

172. T. Ma, N. Collings, and T. Hands, *Exhaust Gas Ignition (EGI)—A New Concept for Rapid Light-Off of Automotive Exhaust Catalyst*, SAE 920400, Society of Automotive Engineers, Warrendale, Pa., 1992.

173. N. Collings and co-workers, *A Linear Catalyst Temperature Sensor for Exhaust Gas Ignition (EGI) and In-Board Diagnostics of Misfire and Catalyst Efficiency*, SAE 930938, Society of Automotive Engineers, Warrendale, Pa., 1993.

General References

Alcoa Alumina Handbook R. B. Bird and co-workers, *Transport Phenomena*, John Wiley & Sons, Inc., New York, 1965.

Code of Federal Regulations, Title 40, Part 86, Washington, D.C., July 1, 1992.

K. C. Taylor, in J. R. Anderson and J. R. Boudart, eds., *Catalysis Science and Technology*, Berlin, 1984, Chapt. 2, pp. 119–170.

G. C. Bond and G. Webb, eds, "Catalysis," *Royal Soc. Chem., London*, **6**, (1984).

D. D. Eley, H. Pines, and P. B. Wiesz, eds., *Advances in Catalysis*, Vol. 33, Academic Press, Inc., New York, 1985.

H. Heinman and J. J. Carberry, eds., *Catalysis Reviews—Science and Engineering*, Vol. 26, Marcel-Dekker, New York, 1984.

C. N. Satterfield, *Heterogeneous Catalysis in Practice*, 1st ed., McGraw-Hill Book Co., Inc., New York, 1980.

J. R. Mondt, *J. Eng. Gas Turbines Power* **109**(2), 200–206 (1987).

C. D. Falk and J. J. Mooney, *Three-Way Conversion Catalysts—Effect of Closed Loop Feedback Control and Other Parameters on Catalyst Efficiency*, SAE 800462, Society of Automotive Engineers, Warrendale, PA., 1980.

M. Luomo, P. Loppi, and R. Lylykaugas, *Evolution of High Cell Density Z-Flow Catalyst*, SAE 930940, Society of Automotive Engineers, Warrendale, Pa., 1993.

John J. Mooney
Engelhard Corporation



EXHAUST CONTROL, INDUSTRIAL

Limits for exhaust emissions from industry, transportation, power generation (qv), and other sources are increasingly legislated (see also EXHAUST CONTROL, AUTOMOTIVE) (1,2). One of the principal factors driving research and development in the petroleum (qv) and chemical processing industries in the 1990s is control of industrial exhaust releases. Much of the growth of environmental control technology is expected to come from new or improved products that reduce such air pollutants as carbon monoxide [630-08-0] (qv), CO, volatile organic compounds (VOCs), nitrogen oxides (NO_x, or other hazardous air pollutants (see AIR POLLUTION). The mandates set forth in the 1990 amendments to the Clean Air Act (CAAA) push pollution control methodology well beyond what, as of this writing, is in general practice, stimulating research in many areas associated with exhaust system control (see AIR POLLUTION CONTROL METHODS). In all, these amendments set specific limits for 189 air toxics, as well as control limits for VOCs, nitrogen oxides, and the so-called criteria pollutants. An estimated 40,000 facilities, including establishments as diverse as bakeries and chemical plants are affected by the CAAA (3).

There are 10 potential sources of industrial exhaust pollutants which may be generated in a production facility (4): (1) unreacted raw materials; (2) impurities in the reactants; (3) undesirable by-products; (4) spent auxiliary materials such as catalysts, oils, solvents, etc; (5) off-spec product; (6) maintenance, ie, wastes and materials; (7) exhausts generated during start-up and shutdown; (8) exhausts generated from process upsets and spills; (9) exhausts generated from product and waste handling, sampling, storage, and treatment; and (10) fugitive sources.

Exhaust streams generally fall into two general categories, intrinsic and extrinsic. The intrinsic wastes represent impurities present in the reactants, by-products, co-products, and residues as well as spent materials used as part of the process, ie, sources (1)–(5). These materials must be removed from the system if the process is to continue to operate safely. Extrinsic wastes are generated during operation of the unit, but are more functional in nature. These are generic to the process industries overall and not necessarily inherent to a specific process configuration, ie, sources (6)–(10). Waste generation may occur as a result of unit upsets, selection of auxiliary equipment, fugitive leaks, process shutdown, sample collection and handling, solvent selection, or waste handling practices (see also WASTE, INDUSTRIAL).

Control Strategy Evaluation

There are two broad strategies for reducing volatile organic compound (VOC) emissions from a production facility: (1) altering the design, operation, maintenance, or manufacturing strategy so as to reduce the quantity or toxicity of air emissions produced, or (2) installing after-treatment controls to destroy the pollutants in the generated air emission stream (5). Whether the exhaust stream contains a specific hazardous air pollutant, a VOC, a nitrogen oxide, or carbon monoxide, the best way to control the pollutant is to prevent its formation in the first place. Many technologies are being developed that seek to minimize the generation of undesirable by-products by modifying specific process materials or operating conditions. Whereas process economics or product quality may restrict the general applicability of these approaches, an increased understanding of the mechanisms and conditions by which a pollutant is created is leading to significant breakthroughs in burner design and operation (for nitrogen oxide control), equipment design, maintenance, and operation for fugitive and vent VOC emission control, and product and waste storage and handling design and operation (for VOC emission control).

Once an undesirable material is created, the most widely used approach to exhaust emission control is the application of add-on control devices (6). For organic vapors, these devices can be one of two types, combustion or capture. Applicable combustion devices include thermal incinerators (qv), ie, rotary kilns, liquid injection combusters, fixed hearths, and fluidized-bed combustors; catalytic oxidization devices; flares; or boilers process heaters. Primary applicable capture devices include condensers, adsorbers, and absorbers, although such techniques as precipitation and membrane filtration are finding increased application. A comparison of the primary control alternatives is shown in Table 1 (see also ABSORPTION; ADSORPTION; MEMBRANE TECHNOLOGY).

Table 1. Emission Control Technologies

Technology	Reduction effectiveness	Recovery	Waste generation	Advantages	Disadvantages
activated carbon	90–98% _o	chemical recovery possible with regeneration	spent carbon or regenerant	good for wide variety of VOCs	carbon replacement, regeneration costs, potential for bed fires
adsorption	75–90% ₊	chemical recovery possible through decanting	spent solvent or regenerant	simple operation	not efficient at low concentration
adsorption in wet scrubbers		distillation			
vapor condensation	50–80% _o	chemical recovery possible through decanting	liquid wastes, needs off-gas treatment	simple operation, effective for high VOC concentration	low removals applicability limits to some VOCs, high power costs
thermal oxidation	99% _o	treatment heat recovery	NO _x generation, CO ₂ generation	handles any VOC concentration	high operating costs, capital costs, temperatures, and maintenance
catalytic oxidation	95–99% _o	heat recovery	spent catalyst regeneration acids and alkalines	simple systems, lower T than thermal economical operation	fouling of catalysts, temperature limits

The most desirable of the control alternatives is capture of the emitted materials followed by recycle back into the process. However, the removal efficiencies of the capture techniques generally depend strongly on the physical and chemical characteristics of the exhaust gas and the pollutants considered. Combustion devices are the more commonly applied control devices, because these are capable of a high level of removal efficiencies, ie, destruction for a variety of chemical compounds under a range of conditions. Although installation of emission control devices requires capital expenditures, these may generate useful materials and be net consumers or producers of energy. The selection of an emission control technology is affected by nine interrelated parameters: (1) temperature, T, of the inlet stream to be treated; (2) residence time; (3) process exhaust flow rate; (4) auxiliary fuel needs; (5) optimum energy use; (6) primary chemical composition of exhaust stream; (7) regulations governing destruction requirements; (8) the gas stream's explosive properties or heat of combustion; and (9) impurities in the gas stream. A process flow diagram for the consideration of these parameters is shown in Figure 1. Given the many factors involved, an economic analysis is often needed to select the best control option for a given application (see ECONOMIC EVALUATION).

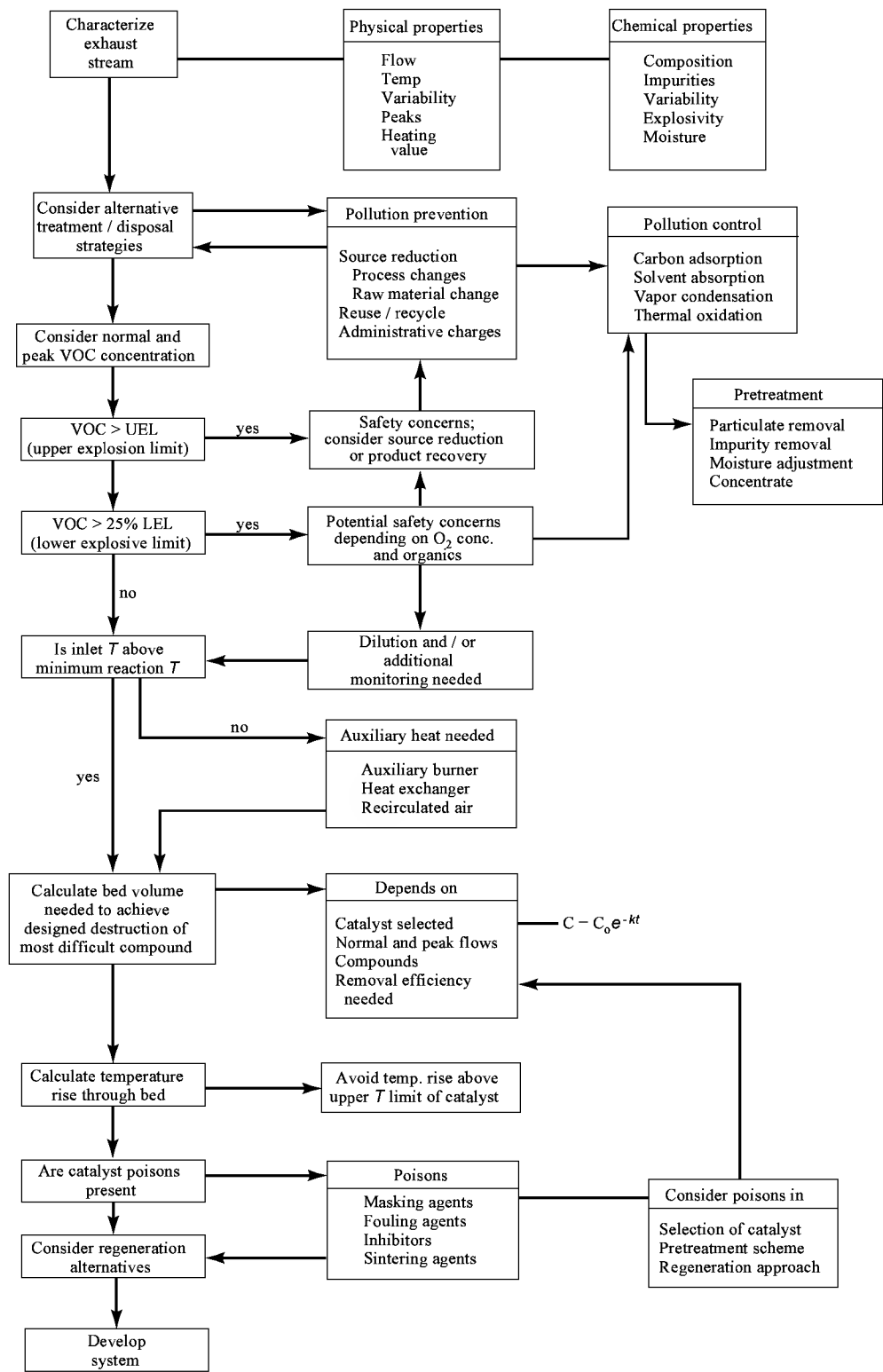


Fig. 1. Process flow diagram for the selection of an exhaust control system.

Capture devices are discussed extensively elsewhere (see AIR POLLUTION CONTROL METHOD). Oxidation devices are either thermal units that use heat alone or catalytic units in which the exhaust gas is passed over a catalyst usually at an elevated temperature. These latter speed oxidation and are able to operate at temperatures well below those of thermal systems.

Oxidization Devices

Thermal Oxidation. Thermal oxidation is one of the best known methods for industrial waste gas disposal. Unlike capture methods, eg, carbon adsorption, thermal oxidation is an ultimate disposal method destroying the objectionable combustible compounds in the waste gas rather than collecting them. There is no solvent or adsorbent of which to dispose or regenerate. On the other hand, there is no product to recover. A primary advantage of thermal oxidation is that virtually any gaseous organic stream can be safely and cleanly incinerated, provided proper engineering design is used (see INCINERATORS) (7).

A thermal oxidizer is a chemical reactor in which the reaction is activated by heat and is characterized by a specific rate of reactant consumption. There are at least two chemical reactants, an oxidizing agent and a reducing agent. The rate of reaction is related both to the nature and to the concentration of reactants, and to the conditions of activation, ie, the temperature (activation), turbulence (mixing of reactants), and time of interaction.

Thermal oxidation relies on a homogenous gas-phase reaction condition. Exhaust emissions from industrial sources usually contain organic compounds (the reducing agents) well mixed with oxygen (the oxidizing agent). Imparting the necessary, uniform temperature for reaction within this mixture is of primary importance in the design of the oxidizer and related equipment. Proper activation requires establishing the minimum required temperature (650–800°C) for an adequate time (0.1–0.3 s). General design consideration is given to minimizing heat input and reactor size under the constraints of time, turbulence, and temperature.

Thermal oxidation devices are widely used, and generally provide a high degree of assurance that the process oxidizes the material in the exhaust gas. The high temperature operation causes other problems, however, especially compared to alternatives such as catalytic oxidation. The thermal oxidation

devices often incur higher fuel costs because of the higher temperatures necessary, and require exotic high temperature materials (see HIGH TEMPERATURE ALLOYS; REFRACTORIES), because the high temperatures entailed can bring about serious mechanical design problems as a result of operating in the temperature range in which metal creep takes place (1,8). In addition, equipment durability is reduced by the extent of thermal cycling. Thermal oxidation systems are susceptible to thermal stress effects which result in distortion of the ductwork and heat-exchange surfaces, creating the potential for cracks and leaks. A further consequence of these high temperatures is that a thermal oxidizer may produce nitrogen oxides (NO_x), and sometimes yield undesirable by-products such as dioxins from chlorinated materials (5).

Some of the problems associated with thermal oxidizers have been attributed to the necessary coupling of the mixing, the reaction chemistry, and the heat release in the burning zone of the system. These limitations can reportedly be avoided by using a packed-bed flameless thermal oxidizer which is under development. This system relies on radiant heat from a large heat sink to raise the temperature of the exhaust gas to its ignition temperature. The heat sink, a ceramic matrix, is preheated radiatively using an electric preheating element, or a natural gas preheater, prior to introducing the exhaust gas. Because the system temperature is reasonably constant, NO_x generation within the flame is minimized. High (99.99%) VOC reductions at low contact times were reported (9). However, as with all thermal oxidation systems, this system is most effective for higher concentration exhaust streams, and requires significant auxiliary heat to treat low concentration streams.

Catalytic Oxidization. A principal technology for control of exhaust gas pollutants is the catalyzed conversion of these substances into innocuous chemical species, such as water and carbon dioxide. This is typically a thermally activated process commonly called catalytic oxidation, and is a proven method for reducing VOC concentrations to the levels mandated by the CAAA (see CATALYSIS). Catalytic oxidation is also used for treatment of industrial exhausts containing halogenated compounds.

As an exhaust control technology, catalytic oxidation enjoys some significant advantages over thermal oxidation. The former often occurs at temperatures that are less than half those required for the latter, consequently saving fuel and maintenance costs. Lower temperatures allow use of exhaust stream heat exchangers of a low grade stainless steel rather than the expensive high temperature alloy steels. Furthermore, these lower temperatures tend to avoid the emissions problems arising from the thermal oxidation processes (10,11).

Critical factors that need to be considered when selecting an oxidation system include (12) (1) waste stream heating value and explosive properties. Low heating values resulting from low VOC concentration make catalytic systems more attractive, because low concentrations increase fuel usage in thermal systems; (2) waste gas components that might affect catalyst performance. Catalyst formulations have overcome many problems owing to contaminants, and a guard bed can be used in catalytic systems to protect the catalyst; (3) the type of fuel available and optimum energy use. Natural gas and no. 2 fuel oil can work well in catalytic systems, although sulfur in the fuel oil may be a problem in some applications (13). Other fuels should be evaluated on a case-by-case basis; and (4) space and weight limitations on the control technology. Catalysts are favored for small, light systems.

There are situations where thermal oxidation may be preferred over catalytic oxidation: for exhaust streams that contain significant amounts of catalyst poisons and/or fouling agents, thermal oxidation may be the only technically feasible control; where extremely high VOC destruction efficiencies of difficult to control VOC species are required, thermal oxidation may attain higher performance; and for relatively rich VOC waste gas streams, ie, having >20–25% lower explosive limit (LEL), the gas stream's explosive properties and the potential for catalyst overheating may require the addition of dilution air to the waste gas stream (12).

Whereas the catalytic converter has been used in automobiles to control air pollutants only since 1975 (5), catalytic oxidation of industrial exhaust emissions began in the late 1940s, and is a reasonably mature technology (14). Initially it was used only in circumstances where an extremely serious odor problem was associated with an industrial system, or where the concentration of organic solvents in the gases to be discharged to the air was high enough that these could be burned and the heat utilized in the process (8). By the mid-1950s there were several dozen catalytic incinerators in California, primarily in Los Angeles county, the first sizable area within the United States to experience a serious air pollution problem. Early applications of this technology involved some serious odor, eye irritation, or visible organic emission problems resulting from halogen poisoning and catalyst fouling (15).

The chemical industry was the first to utilize catalytic oxidation extensively for emission control, building units capable of treating up to $50 \text{ m}^3/\text{s}$ (100,000 scfm) of exhaust gas containing VOCs. Catalytic systems accounted for roughly one-fourth of the \$200 million market for VOC control systems in 1992, and over one thousand catalytic oxidization devices were in place by the end of that year (5).

Catalysts. A catalyst has been defined as a substance that increases the rate at which a chemical reaction approaches equilibrium without becoming permanently involved in the reaction (16). Thus a catalyst accelerates the kinetics of the reaction by lowering the reaction's activation energy (5), ie, by introducing a less difficult path for the reactants to follow. For VOC oxidation, a catalyst decreases the temperature, or time required for oxidation, and hence also decreases the capital, maintenance, and operating costs of the system (see CATALYSIS).

A key feature of a catalyst is that the catalytic material is not consumed by the chemical oxidation reactions, rather it remains unaltered by the reactions which occur generally on its surface and thus remains available for an infinite number of successive oxidation reactions.

Many chemical elements exhibit catalytic activity (5) which, within limits, is inversely related to the strength of chemisorption of the VOCs and oxygen, provided that adsorption is sufficiently strong to achieve a high surface coverage (17). If the chemisorption is too strong, the catalyst is quickly deactivated as the active sites become irreversibly covered. If the chemisorption is too weak, only a small fraction of the surface is covered and the activity is very low (17) (Fig. 2).

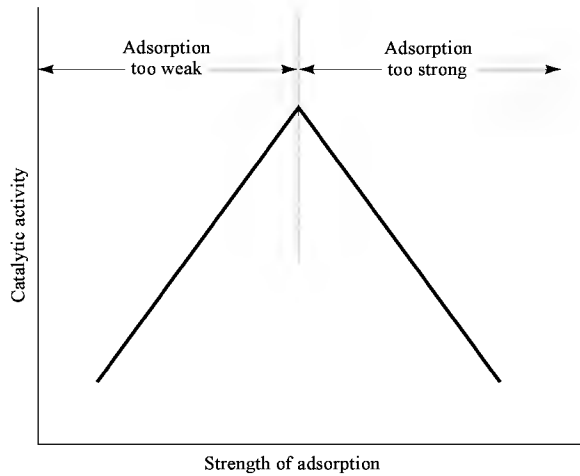


Fig. 2. Catalytic activity as a function of adsorption strength (17).

Courtesy of Oxford University Press.

Catalysts vary both in terms of compositional material and physical structure (18). The catalyst basically consists of the catalyst itself, which is a finely divided metal (14,17,19); a high surface area carrier; and a support structure (see CATALYSTS, SUPPORTED). Three types of conventional metal catalysts are used for oxidation reactions: single- or mixed-metal oxides, noble (precious) metals, or a combination of the two (19).

The precious metal or metal oxide imparts high intrinsic activity, the carrier provides a stable, high surface area for catalyst dispersion, and the mechanical support gives a high geometric surface area for physical support and engineering design features (20). Only the correct combination of these

components provides suitable performance and long catalyst life of a properly designed catalytic system (21).

Metal Oxides. The metal oxides are defined as oxides of the metals occurring in Groups 3–12 (IIIB to IIB) of the Periodic Table. These oxides, characterized by high electron mobility and the positive oxidation state of the metal, are generally less active as catalysts than are the supported noble metals, but the oxides are somewhat more resistant to poisoning. The most active single-metal oxide catalysts for complete oxidation of a variety of oxidation reactions are usually found to be the oxides of the first-row transition metals, V, Cr, Mn, Fe, Co, Ni, and Cu.

Noble Metals. Noble or precious metals, ie, Pt, Pd, Ag, and Au, are frequently alloyed with the closely related metals, Ru, Rh, Os, and Ir (see PLATINUM GROUP METALS). These are usually supported on a metal oxide such as α -alumina, α -Al₂O₃, or silica, SiO₂. The most frequently used precious metal components are platinum [7440-06-4], Pt, palladium [7440-05-3], Pd, and rhodium [7440-16-6], Rh. The precious metals are more commonly used because of the ability to operate at lower temperatures. As a general rule, platinum is more active for the oxidation of paraffinic hydrocarbons; palladium is more active for the oxidation of unsaturated hydrocarbons and CO (19).

Each precious metal or base metal oxide has unique characteristics, and the correct metal or combination of metals must be selected for each exhaust control application. The metal loading of the supported metal oxide catalysts is typically much greater than for noble metals, because of the lower inherent activity per exposed atom of catalyst. This higher overall metal loading, however, can make the system more tolerant of catalyst poisons. Some compounds can quickly poison the limited sites available on the noble metal catalysts (19).

Carrier. The metal catalyst is generally dispersed on a high surface area carrier, ie, the carrier is given a washcoat of catalyst, such that very small (2–3 nm dia) precious metal crystallites are widely dispersed over the surface area, serving two basic functions. It maximizes the use of the costly precious metal, and provides a large surface area thereby increasing gas contact and associated catalytic reactions (18).

Proper selection of the carrier is critical (20). For example, in most cases, the carrier of a precious metal catalyst is a high surface area alumina having an effective surface area in the order of 120 m² g of material. Alumina is often used because of its unique phase transformation properties. Various phases of the aluminum hydroxides exist as a function of temperature and starting phase. For a catalyst to be stable, the correct starting phase of alumina must be selected for the projected commercial operating temperature. Otherwise the alumina may undergo a transition during operation resulting in a carrier of less surface area, and hence less catalytic activity (1) (see ALUMINUM COMPOUNDS, ALUMINUM OXIDES).

Efforts to redesign catalyst formulations involve both catalyst and washcoat. One thrust of this research is to manipulate the catalytic surface so that it can handle larger quantities of catalyst poisons and to incorporate more catalytic sites, redistributed within the washcoat to make them more accessible to exhaust molecules. Altering the composition of the alumina washcoat by including various nonprecious metal oxides, such as oxides of barium, cerium, and lanthanum as stabilizers, is being looked at to promote catalyst activity before sintering and stabilize precious metal dispersion (18). Reformulation efforts are aided by use of computer controls and cleaner reactants and continuous monitoring, all of which help make exhaust composition more predictable (1).

The Support Structure. After the catalytic element is placed on the high surface area carrier, it is deposited on a mechanical support structure which determines the form of the catalyst. The support structure may have many forms, such as spheres, pellets, woven mesh, screen, honeycomb, or other ceramic matrix structures designed to maximize catalyst surface area (6). Some of the advantages of these different supports are listed in Table 2.

Table 2. Conventional Catalyst Bed Geometries^a

Geometry	Advantages	Disadvantages
metal ribbons	low pressure drop; high surface-to-volume	less active than ceramic-supported catalysts
spherical pellets	can be used in both fixed and fluidized beds	high pressure drop; attrition problem
ceramic rods	low pressure drop	low surface-to-volume ratio
ceramic honeycomb	low pressure drop; high surface-to-volume ratio	may have nonuniform catalyst coating
metal honeycomb	low pressure drop; high surface-to-volume ratio; high mechanical strength	less active than ceramic-supported catalysts

^a Ref. 14.

The pelleted and honeycomb support structures are most widely used; honeycombs are the most commonly employed. Pelleted structures are generally spherical beads or cylinders having diameters ranging from 0.16–0.64 cm. The pellets are assembled into a packed bed containing large numbers of these pellets through which the exhaust passes. The honeycomb supports are monolithic structures having numerous parallel channels through which the exhaust passes, the channel sizes ranging from about 8 to 50 cells cm² of catalyst frontal area. Each cell has a width opening ranging from 0.29 to 0.13 cm, respectively. Some commercial honeycombs are available from 1.6 to 100 cells cm² (20). The shape in the individual honeycomb channel is unlimited, eg, circle, square, triangle, etc.

Although more expensive to fabricate than the pelleted catalyst, and usually more difficult to replace or regenerate, the honeycomb catalyst is more widely used because it affords lower pressure losses from gas flow; it is less likely to collect particulates (fixed-bed) or has no losses of catalyst through attrition, compared to fluidized-bed; and it allows a more versatile catalyst bed design (18), having a well-defined flow pattern (no channeling) and a reactor that can be oriented in any direction.

The honeycomb structure is either fabricated of ceramic or stainless steel. The high surface area carrier and catalytic precious metal crystallites are coated onto the walls of the channels in the honeycomb. The honeycomb catalyst blocks generally range form 15 to 30 cm square at depths from 5 to 10 cm. These blocks are packed into larger modules containing many catalyst blocks. Flow through a honeycomb catalyst structure is shown in Figure 3 (18,22). A typical 30 cells cm² honeycomb structure has about 4600 m² of geometric wall area per cubic meter of catalyst volume (18). The actual shape of the individual honeycomb channel is unlimited, eg, it may be circular, square, triangular, etc. In addition, the channel density can be varied. Commercial honeycombs are available that range from 1.5 to 100 cells cm² (10 to 600 cells in.²) (20).

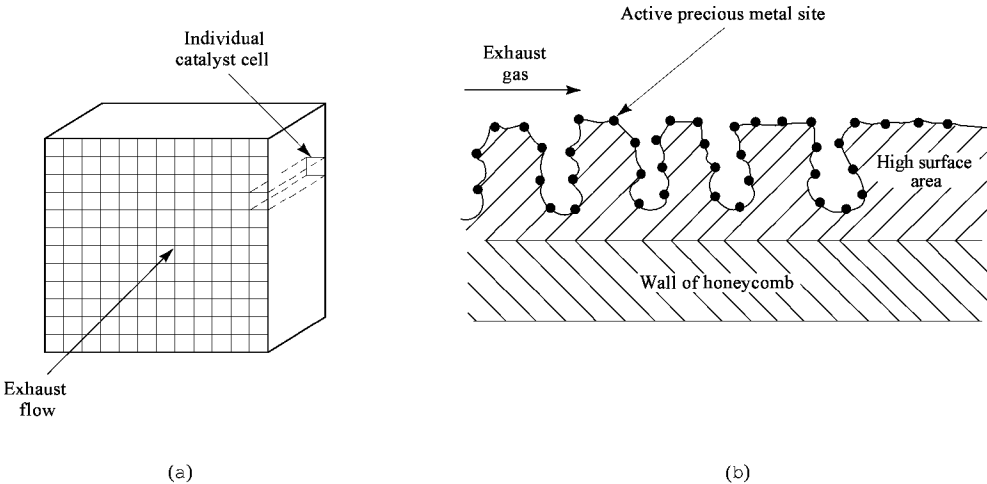


Fig. 3. Schematic of the flow through (a) a honeycomb catalyst structure and (b) a cross section of a honeycomb channel.

Only by using the carrier can the catalyst be sufficiently active because the majority of applications require 10 to 100 m² g of surface area (20). Surface areas for a typical monolith support structure and a carrier are given (20).

Catalyst	BET surface area, m ² g
geometric smooth monolith wall	0.003
natural porosity of monolith wall	0.3
carrier on monolith wall	20

Mechanistic Models. A general theory of the mechanism for the complete heterogeneous catalytic oxidation of low molecular weight vapors at trace concentrations in air does not exist. As with many catalytic reactions, however, certain observations have led to a general hypothesis (17).

The overall process of any catalytic reaction is a combination of mass-transfer, describing transport of reactants and products to and from the interior of a solid catalyst, and chemical reaction kinetics, describing chemical reaction sequences on the catalyst surface. The most cost-effective catalytic oxidation systems require use of a solid catalyst material having a high specific surface area, ie, high surface area per net weight of catalyst. The presence of many small pores necessarily introduces pore transport diffusion resistance as a factor in the overall, or global, kinetics. The overall process consists of (17,23): (1) transport of reactants from the bulk fluid through the gas film boundary layer to the surface of the catalytic particle; (2) transport of reactants into the catalyst particle by diffusion through the catalyst pores; (3) chemisorption of at least one reactant on the catalyst surface; (4) chemical reaction between chemisorbed species or between a chemisorbed species and a physisorbed or fluid-phase reactant; (5) desorption of reaction products from the catalyst surface; (6) diffusive transport of products through the catalyst pores to the surface of the catalyst particle; and (7) mass transfer of products through the exterior gas film to the bulk fluid.

In principle, any of these steps or some combination can be rate controlling. In practice, temperature plays a primary role in determining the rate-controlling stage. Any comprehensive analysis of actual catalytic oxidation systems of practical interest must include a quantitative understanding of the relative effects of mass transfer (steps 1,2,6,7) and surface reaction (steps 3,4,5). The temperature relationship of these two mechanisms is shown in Figure 4 (17,18,20). As a catalyst is heated, conversion of the pollutant is negligible until a critical temperature is reached, then the rate of conversion increases rapidly with rising temperature. This is referred to as the kinetically limited region. Conversion increases in this region because catalytic reaction rates increase with temperature, until the catalyst's normal operating temperature is achieved. Then the conversion rate increases only slightly with further temperature rise in the mass-transfer limited region. At some advanced temperature, the conditions reach a point where thermal oxidation begins to play a role, and the rate of conversion again increases rapidly.

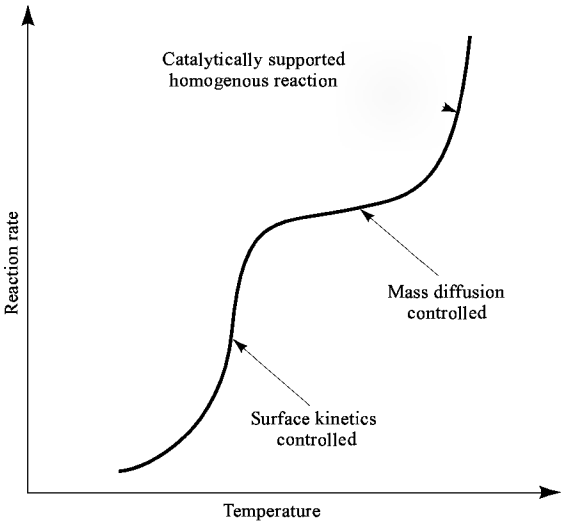


Fig. 4. Reaction rate profile as a function of temperature (20).

In the mass-transfer limited region, conversion is most commonly increased by using more catalyst volume or by increasing cell density, which increases the catalytic wall area per volume of catalyst. When the temperature reaches a point where thermal oxidation begins to play a role, catalyst deactivation may become a concern.

Reaction Rate. The kinetics for a single catalytic reaction can be modeled as

$$r_m = k(T) f(C) \eta$$

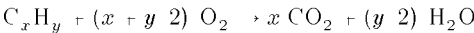
where r_m is the rate of the main reaction; $k(T)$ is the rate constant, a function of temperature, T ; $f(C)$ is the function of reactant and product concentration, C ; and η is the effectiveness factor, which accounts for pore-diffusional resistance (24). The form of the terms $k(T)$ and $f(C)$ depends on the kinetic model for the system. Kinetic models for the catalytic oxidation can either be empirical or mechanistic.

Empirical Models. In the case of an empirical equation, the model is a power law rate equation that expresses the rate as a product of a rate constant and the reactant concentrations raised to a power (17), such as

$$r_m = k C_1^a C_2^b$$

where r_m is the rate constant; C_1 is the concentration of reactant 1; C_2 is the concentration of reactant 2; and a and b are empirically determined reaction orders.

For combustion of simple hydrocarbons, the oxidation reactions appear to follow classical first-order reaction kinetics sufficiently closely that practical designs can be established by application of the empirical theory (8). For example, the general reaction for a hydrocarbon:



can be represented by the rate equation

$$r_m = (dC/dt) = -kC$$

where C = hydrocarbon concentration, r_m = rate of change of contaminant concentration, t = time, and k = reaction rate constant, which must be determined experimentally from the burning of various organic materials. The pattern of variation with t is predictable from kinetic theory and follows the

Arrhenius equation,

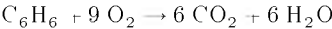
$$k = A\exp(-\Delta E/RT)$$

where A is the Arrhenius collision constant, ΔE the activation energy, and R is the universal gas constant. A catalyst increases the rate of reaction by adsorbing gas molecules on catalytically active sites. The catalyst may function simply to bring about a higher concentration of reactive materials at the surface than is present in the bulk gas phase, which has the effect of increasing the collision constant, A , or the catalyst may modify a molecule of adsorbed gas by adding or removing an electron or by physically opening a bond. This has the effect of decreasing the activation energy, ΔE , in the Arrhenius equation. In either circumstance, it is necessary for the reactive materials to reach the active catalyst surface by diffusion through the gas phase, and for the reaction products to leave the surface. For the conditions encountered in most hydrocarbon emission control applications, the oxygen partial pressure is much larger than the organic reactant partial pressure, and can be treated as a constant.

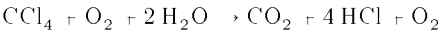
Mechanistic kinetic expressions are often used to represent the rate data obtained in laboratory studies, and to explain quantitatively the effects observed in the field. Several types of mechanisms have been proposed. These differ primarily in complexity, and on whether the mechanism assumes that one compound that is adsorbed on the catalyst surface reacts with the other compound in the gas phase, eg, the Eley-Rideal mechanism (23); or that both compounds are adsorbed on the catalyst surface before they react, eg, the Langmuir-Hinshelwood mechanism (25).

The volatile organic compounds on the list of hazardous air pollutants under the CAAA have been classified into four main categories: (1) pure hydrocarbons (qv), (2) halogenated hydrocarbons (see CHLOROCARBONS AND CHLOROHYDROCARBONS), (3) nitrogenated hydrocarbons (see CYANIDES), and (4) oxygenated hydrocarbons (see ALDEHYDES; ETHERS; KETONES). The compounds in these groups are characterized by the following oxidation reactions (26):

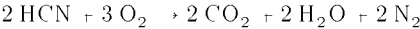
Hydrocarbons



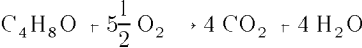
Halogenated hydrocarbons



Nitrogenated hydrocarbons



Oxygenated hydrocarbons



Temperature reaction rate profiles for representatives compounds are available (21,26). Particularly important are the operating temperatures required before destruction is initiated. Chemical reactivity by compound class from high to low is (27) alcohols $\succ$ cellsolves $\succ$ dioxane $\succ$ aldehydes $\succ$ aromatics $\succ$ ketones $\succ$ acetates $\succ$ alkanes $\succ$ chlorinated hydrocarbons. In general, within a class the higher the molecular weight, the higher the relative destructibility. All of these compound classes, except chlorinated hydrocarbons, can be destroyed with 98–99% efficiency at sufficiently low space velocities and $\geq$ or high enough inlet temperatures (28). Table 3 (22) presents oxidation temperatures for a number of hydrocarbons.

Table 3. Ignition Temperatures for 90% Conversion

Component	Temp, °C
hydrogen	93
acetylene	177
carbon monoxide	218
cyclohexanone	218
propylene	232
toluene	232
2-propanol	260
ethylene	260
benzene	260
xylene	260
ethanol	260
methyl ethyl ketone	274
ethyl acetate	288
cyclohexane	288
<i>n</i> -hexane	316
methyl isobutyl ketone	316
propane	399
methane	427

^a Ref. 22.

Historically, the destruction efficiency for chlorinated hydrocarbons is quite low. In addition, tests conducted after the chlorinated hydrocarbon is treated show that the catalyst is partially deactivated. More recent advancements in catalyst technology have resulted in the development of a number of catalysts and catalytic systems capable of handling most chlorinated hydrocarbons under a variety of conditions (19).

Mixture Effects. Care must be taken in determining the oxidation kinetics for a mixture of chemicals (29). In principle, given one set of conditions and a two-component mixture, the overall conversion of one component A may be controlled by mass transfer to the catalyst surface and the conversion of another component B by surface-reaction kinetics. Of course, the controlling regime (mass transfer or reaction) can change with temperature. Thus for two independent parallel reactions the combined effect of diffusional and reaction rate resistances can have a considerable influence on the relative rate of the two reactions. Additionally, a third, fourth, or n th component can conceivably affect the other components by, for instance, competing more successfully for active surface sites than B while simultaneously influencing the mass transfer of A. Thus even for a simple two-or three-component mixture, interpretation of observed results can be difficult. Extrapolation of mixture behavior from single-component data is ill-advised.

In a mixture of n -hexane and benzene (29), the deep catalytic oxidation rates of benzene and n -hexane in the binary mixture are lower than when these compounds are singly present. The kinetics of the individual compounds can be adequately represented by the Mars-VanKrevelen mechanism. This model needs refinements to predict the kinetics for the mixture.

One important consideration in any catalyst oxidation process for a complex mixture in the exhaust stream is the possible formation of hazardous incomplete oxidation products. Whereas the concentration in the effluent may be reduced to acceptable levels by mild basic aqueous scrubbing or additional vent gas treatment, studying the kinetics of the mixture and optimizing the destruction cycle can drastically reduce the potential for such emissions.

Design and Operation. The destruction efficiency of a catalytic oxidation system is determined by the system design. It is impossible to predict *a priori* the temperature and residence time needed to obtain a given level of conversion of a mixture in a catalytic oxidation system. Control efficiency is determined by process characteristics such as concentration of VOCs emitted, flow rate, process fluctuations that may occur in flow rate, temperature, concentrations of other materials in the process stream, and the governing permit regulation, such as the mass-emission limit. Design and operational characteristics that can affect the destruction efficiency include inlet temperature to the catalyst bed, volume of catalyst, and quantity and type of noble metal or metal oxide used.

Catalytic oxidation systems are normally designed for destruction efficiencies that range from 90 to 98% (27). In the early 1980s, typical design requirements were for 90% or higher VOC conversions. More recently, however, an increasing number of applications require 95 to 98% conversions to meet the more stringent emission standards (20).

Operational Considerations. The performance of catalytic incinerators (28) is affected by catalyst inlet temperature, space velocity, superficial gas velocity (at the catalyst inlet), bed geometry, species present and concentration, mixture composition, and waste contaminants. Catalyst inlet temperatures strongly affect destruction efficiency. Mixture compositions, air-to-gas (fuel) ratio, space velocity, and inlet concentration all show marginal or statistically insignificant effects (30).

Operating Temperature. The operating temperature needed to achieve a particular VOC destruction efficiency depends primarily on the species of pollutants contained in the waste stream, the concentration of the pollutants, and the catalyst type (14). One of the most important factors is the hydrocarbon species. Each has a catalytic initiation temperature which is also dependent on the type of catalyst used (14).

For a given inlet temperature, the quantity of supplied heat may be provided by (6) the heat supplied from the combustion of supplemental fuel, the sensible heat contained in the emission stream as it enters the catalytic system, and the sensible heat gained by the emission stream through heat exchange with hot flue gas (6). Three types of systems for catalytic oxidation of VOCs are shown in Figure 5 (11). The simplest (Fig. 5a) uses a direct contact open flame to preheat the gas stream upstream of the catalyst. The second (Fig. 5b) involves only a catalyst bed over which the gas stream passes, usually after some indirect preheating. The third (Fig. 5c) involves more extensive indirect preheating and heat exchange. The difference in the three configurations is the method for preheating the gas.

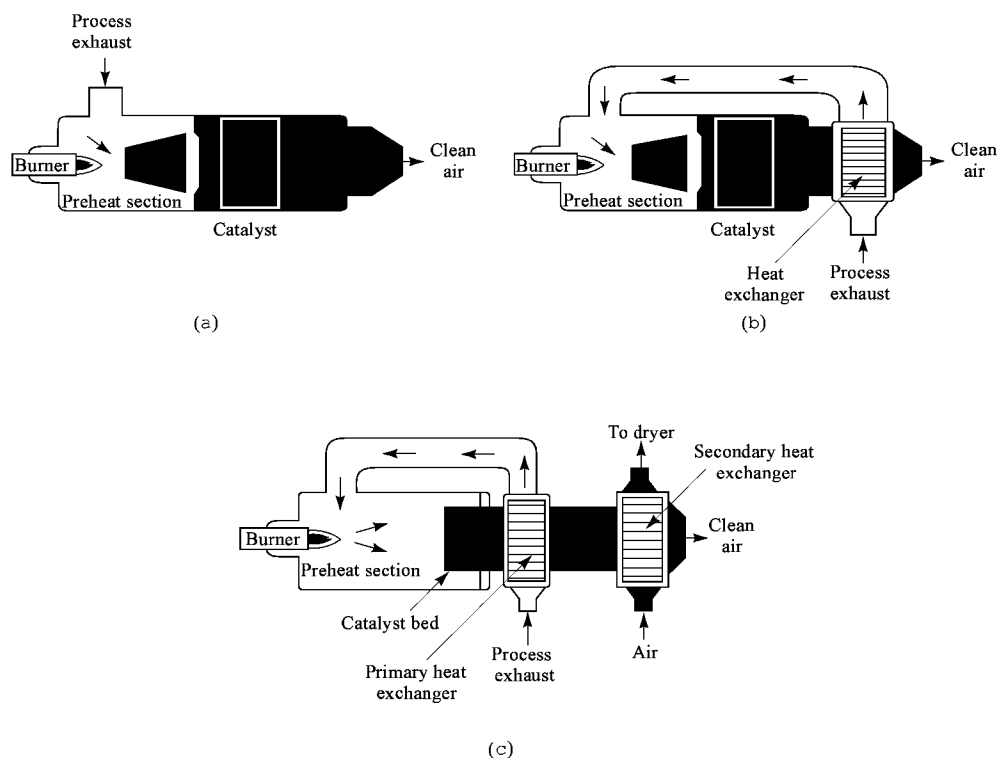


Fig. 5. Catalytic system designs (11) of (a) basic VOC catalytic converter containing a preheater section, a reactor housing the catalyst, and essential controls, ducting, instrumentation, and other elements; (b) a heat exchanger using the cleaned air exiting the reactor to raise the temperature of the incoming process exhaust; and (c) extracting additional heat from the exit gases by a secondary heat exchanger.

There are two general temperature policies: increasing the temperature over time to compensate for loss of catalyst activity, or operating at the maximum allowable temperature. These temperature approaches tend to maximize destruction, yet may also lead to loss of product selectivity. Selectivity typically decreases with increasing temperature; faster deactivation; and increased costs for reactor materials, fabrication, and temperature controls.

Reactor Design. The catalytic reactor is designed to be operated in the mass-transfer controlled catalytic region. The prime design parameter is the geometric surface area. The honeycomb catalyst shows substantial advantage over other forms because of the high geometric surface areas obtainable with low pressure drop (20).

Catalyst Selection. The choice of catalyst is one of the most important design decisions. Selection is usually based on activity, selectivity, stability, mechanical strength, and cost (31). Stability and mechanical strength, which make for steady, long-term performance, are the key characteristics. The basic strategy in process design is to minimize catalyst deactivation, while optimizing pollutant destruction.

Both catalyst space velocity and bed geometry play a role. The gas hourly space velocity (GHSV) is used to relate the volumetric flow rate to the catalyst volume. GHSV has units of inverse hour and is defined as the volume flow rate per catalyst volume.

The size of the catalyst bed depends mainly on the degree of VOC reduction required (14). VOC destruction efficiencies up to 95% can usually be attained using reasonable space velocities (14). However, the low GHSVs, and subsequently high catalyst volumes required to achieve extremely high (eg, 99%) conversions, can sometimes make catalytic oxidation uneconomical. Conventional bed geometries may be found in the literature (14).

Process Conditions. To effectively design a catalytic control system, the Manufacturers of Emissions Controls Association recommends the following data be obtained (5): list of all VOCs present and range of concentration of each, flow rate of exhaust and expected variability, oxygen concentration in exhaust and expected variability, temperature of exhaust and expected variability, static pressure, potential uses for heat recovery, particular performance criteria and or regulations to be met, capture efficiency, ie, fraction of all organic vapors generated by the processes that are directed to the control device, presence of hydrocarbon aerosols in the effluent exhaust, identity and quantity of all inorganic and organic particulate, amount of noncombustibles, presence of possible catalyst deactivators, and anticipated start-up/shutdown frequency of the system.

Pilot Studies. Applications requiring the reduction of VOC emissions have increased dramatically. On-site pilot tests are beneficial in providing useful information regarding VOC emission reduction applications. Information that can be obtained includes optimum catalyst operating conditions, the presence of contaminants in the gas stream, and the effects of these contaminants (see PILOT PLANTS AND MICROPLANTS).

Catalyst Inhibition. A number of potential applications for catalytic oxidation of organic materials have resulted in serious odor or eye

irritation, or visible emission problems (8). Some of these failures are a result of fouling of the catalyst surface. Others occur because materials such as halogens in the gas stream interfere with or suppress the activity of the catalyst, or because the substances react with the precious metals, rendering them permanently inactive. Finally, all catalysts eventually deteriorate by aging or thermal processes (8).

Many of the exhaust streams that must be purified contain significant amounts of halogenated organics, such as polychlorinated ethanes and ethylenes vented in the manufacture of vinyl chloride monomer or released in usage solvents (32) (see VINYL POLYMERS; SOLVENTS, INDUSTRIAL). However, the catalysts used in the conventional catalytic oxidation are severely inhibited by the halogen atoms in these compounds (32). Other trace contaminants of concern in air streams may include phosphorus, nitrogen, and sulfur-containing compounds. Whereas gases containing chlorine, sulfur, and other atoms can deactivate supported noble metal catalysts such as platinum, chlorinated VOC can be treated by certain supported metal oxide catalysts (7).

The four basic mechanisms of catalyst decay are shown in Figure 6 (5,18,24). These are fouling or masking, poisoning, thermal degradation through aging or sintering, and loss of catalyst material through formation and escape of vapors. Poisoning and vapor transport are basically chemical phenomena, whereas fouling is mechanical. Table 4 lists substances that inhibit catalyst activity (5).

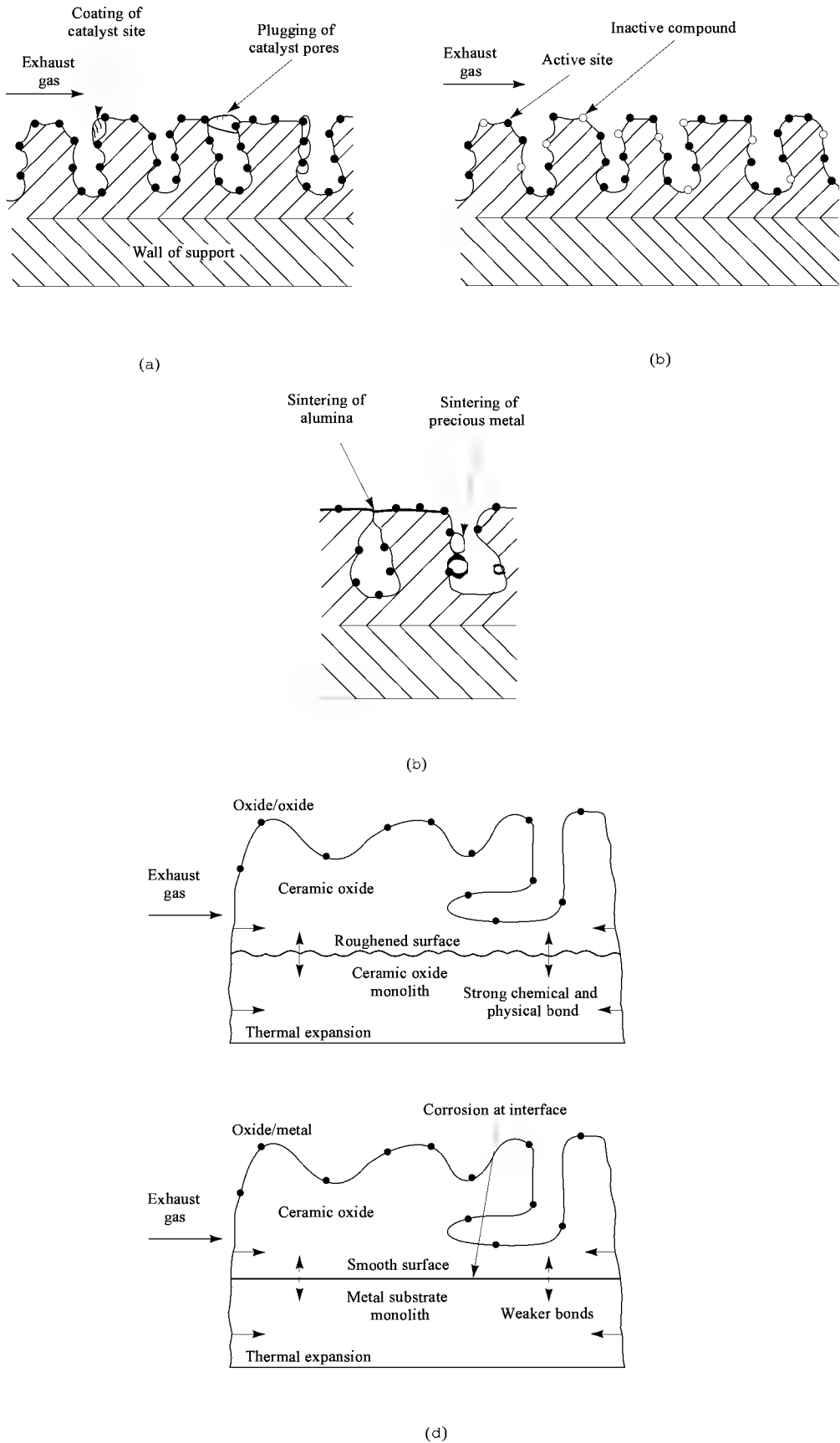


Fig. 6. Catalyst inhibition mechanisms where (●) are active catalyst sites; ▨, the catalyst carrier; and ▩ the catalytic support: (a) masking of catalyst; (b) poisoning of catalyst; (c) thermal aging of catalyst; and (d) attrition of ceramic oxide metal substrate monolith system, which causes the loss of active catalytic material resulting in less catalyst in the reactor unit and eventual loss in performance.

Table 4. Substances that Inhibit Catalyst Activity^a

Type of inhibitor	Effect	Examples	Regeneration
fast-acting inhibitors	reduction of catalyst activity at rate depending on concentration and temperature	phosphorus, bismuth, lead, arsenic, antimony, mercury	catalyst regeneration is sometimes difficult or impossible
slow-acting inhibitors	reduction of catalyst activity; higher concentrations than those of fast-acting catalyst inhibitors may be tolerated	iron, tin, silicon	catalyst regeneration remains difficult or impossible
reversible inhibitors maskers	surface coating of catalyst active area; rate also dependent on concentration and temperature	sulfur, halogens, silicon, zinc, phosphorus	regeneration is possible
surface maskers	surface coating of catalyst active areas	organic solids	removed by increasing catalyst temperature or by acid and alkaline washing
surface eroders and maskers	surface coating of catalyst active area, or erosion of catalyst surface; both result in loss of catalyst activity; rate dependent on particle size, grain loading, and gas stream velocity	inert particulates	surface coating is easily removed by washing
thermal degradation and sintering	loss of catalyst surface area because of catalyst dispersion and crystal growth, or catalyst support collapse through sintering	higher temperature for extended time, temperature excursions, hot spots in bed	regeneration generally very difficult; best avoided by operating in optimum temperature range and avoiding temperature excursions
vapor transport and attrition	loss of catalytic material through formation of metal carbonyl oxides, sulfides, and halides, or surface shear effects resulting from exhaust gas velocity, particulates, or thermal shock	CO, NO, hydrogen sulfide, halogens, and particulates	must replace lost catalytic material; vaporization generally not a factor; attrition particularly important in fluidized beds

^a Refs. 5 and 22.

Masking or Fouling. Masking or fouling is a physical deposition of species from the fluid phase onto the catalyst surface (Fig. 6a) that results in blockage of reaction sites or pores (24). Masking or fouling is caused by a gradual accumulation of noncombusted, solid material that mechanically coats the catalyst's surface and prevents or slows down the diffusion of reactants to the catalyst.

Typical masking or fouling agents include (5,8,21,24) airborne dust or dirt; metal oxides formed from materials in the process, such as silicon dioxide ash remaining when silicone compounds are oxidized; aggregate compound formation on the catalyst surface, ie, phosphorus for lubricating oils; corrosion products from the duct system; and organic char or tars formation from incomplete combustion products, often caused by too low a reactor operating temperature.

Low levels of particulates or potential poisons can sometimes be tolerated without a dramatic decrease in performance. Generally it has been recommended that the maximum particulate concentration not exceed 115 mg m⁻² and that the maximum poison concentration not exceed 25 ppm (14). In addition, every effort should be made to avoid flow over a cold catalyst bed for any extended period of time, as a process stream containing volatile organics may condense on a cold catalyst bed (5).

Combustible masking materials such as organic char may be partially or completely removed by periodic elevations of the catalyst bed temperature. Noncombustible masking materials may be removed by air lancing or aqueous washing generally with a leaching solution (20,21).

Poisons. Halogens, sulfur dioxide [7446-09-5], SO₂, nitrogen dioxide [10102-44-0], NO₂, and numbered other materials act as catalyst suppressants for precious-metal oxidation catalysts. These compounds tend to adsorb strongly on the catalytic surface, preventing the reactants from doing so. The strength of adsorption is ordinarily such that the suppressant materials can gradually be stripped off after there is no longer a concentration of suppressant materials in the gas stream passing through the catalyst (8). In other cases, the adsorption is irreversible. A poison blocks the catalytic sites, and may also induce changes in the surface to result in formation of compounds (24). Active precious-metal sites become inactive, reducing catalyst performance (see Fig. 6b).

At low (<450°C) temperatures, the presence of these materials, particularly the oxides, leads to simple masking or fouling. In some cases, a catalyst that shows reduced activity believed to be from poisoning may simply be masked, and activity can be rejuvenated by cleaning with aqueous leaching solutions (21).

Poisoning is operationally defined. Often catalysts believed to be permanently poisoned can be regenerated (5) (see CATALYSTS, REGENERATION). A species may be a poison in some reactions, but not in others, depending on its adsorption strength relative to that of other species competing for catalytic sites (24), and the temperature of the system. Catalysis poisons have been classified according to chemical species, types of reactions poisoned, and selectivity for active catalyst sites (24).

Group 14 and 15 (VA and VIA) elements act as poisons. The interaction depends on the poison's oxidation state and chemical structure. For sulfur, the order of decreasing poisoning activity is H₂S > SO₂ > SO₂²⁻. Adsorption studies indicate that H₂S adsorbs strongly and dissociates on nickel surfaces. The sulfur adsorbs essentially irreversibly and over most of the catalyst–metal surface. It has been observed that SO₂ and SO₃ also poison catalysts differently (13); SO₂ selectively adsorbs on Pt or Pd in an oxidation catalyst, whereas SO₃ reacts with the Al₂O₃ carrier, forming Al₂(SO₄)₃, which destroys the structure of the catalyst. The latter can be prevented by using a more inert support such as SiO₂ or TiO₂. The former requires a change in operating conditions such as a higher temperature. When No. 2 fuel oil is used in some gas turbines, the sulfur compound in the fuel oil can be converted to SO₂ at levels of 40 to 150 ppm in the exhaust. For such applications, the presence of 100 to 200 ppm SO₂ can require 150 to 200°C higher temperature for the catalyst to give the same CO conversion as without SO₂.

Toxic heavy metals and ions, eg, Pb, Hg, Bi, Sn, Zn, Cd, Cu, and Fe, may form alloys with catalytic metals (24). Materials such as metallic lead, zinc, and arsenic react irreversibly with precious metals and make the surface unavailable for catalytic reactions. Poisoning by heavy metals ordinarily destroys the activity of a precious-metal catalyst (8).

Molecules having unsaturated bonds, eg, CO, NO, HCN, and benzene, may chemisorb through multiple bonds (24).

Catalysts having improved poison resistance have been developed. Catalysts are available that can destroy chlorine-, fluorine-, or bromine-containing organic compounds (5).

Thermal Degradation and Sintering. Thermally induced deactivation of catalysts may result from redispersion, ie, loss of catalytic surface area because of crystal growth in the catalyst phase (21,24,33) or from sintering, ie, loss of catalyst-support area because of support collapse (18). Sintering processes generally take

place at high ($>500^{\circ}\text{C}$) temperatures and are generally accelerated by the presence of water vapor (see Fig. 6c). Another thermal effect is the transformation of catalytic phases to noncatalytic ones, eg, the reaction of nickel and alumina to form nickel aluminate (24). Each catalyst has a recommended temperature window of operation. At temperatures above this window (usually $>760^{\circ}\text{C}$), sintering can occur.

Loss of Catalyst by Vapor Transport. The direct volatilization of catalytic metals is generally not a factor in catalytic processes, but catalytic metal can be lost through formation of metal carbonyl oxides, sulfides, and halides in environments containing CO , NO , O_2 and H_2S , and halogens (24).

The ceramic oxide carrier is bonded to the monolith by both chemical and physical means. The bonding differs for a ceramic monolith and a metallic monolith. Attrition is a physical loss of the carrier from the monolith from the surface shear effects caused by the exhaust gas, a sudden start-up or shutdown causing a thermal shock as a result of different coefficients of thermal expansion at the boundary between the carrier and the monolith, physical vibration of the catalyzed honeycomb, or abrasion from particulates in the exhaust air (21) (see Fig. 6d).

Avoiding Catalyst Deactivation. Catalyst deactivation is more easily prevented than cured. Poisoning by impurities may be prevented by removing impurities from the reactants. Carbon deposition and coking may be prevented by minimizing formation of precursors and manipulating mass-transfer regimes so as to minimize the carbon's or coke's effect on activity. Most sintering is irreversible, or reversible only with great difficulty, so it is important to choose reaction, ie, lower temperatures, that do not sinter the catalyst. Additionally, when process upsets that could release inhibitors or cause small fluctuations in the heating value of the oxidizer are highly probable a thermal system is favored over a catalytic one (7).

Except for No. 2, fuel oil should not be considered as auxiliary fuel when using a catalytic system because of the sulfur and vanadium the fuel oil may contain (7). In some cases even the sulfur in No. 2 fuel oil can present a problem. Galvanized metal should not be used in process ovens or ductwork because zinc is a catalyst poison.

Proper system design and catalyst maintenance are key to minimizing deactivation and providing long-term catalyst service. For example, control of air dilution, use of temperature control loops, and use of catalysts having high intrinsic thermal stability can provide necessary protection against high temperature damage caused by reaction exotherms and from operational upsets (20).

Experimental Evaluation. Often the deactivation kinetics for a catalytic oxidation system can be evaluated in a series of laboratory studies (24). Reactors should be gradientless with respect to reactant poison concentration and temperature. Heat- and mass-transfer effects should be avoided because these disguise the intrinsic kinetics. Experiments should be designed to study one deactivation process at a time, and accelerated targets must be representative of the process. Deactivation can be accelerated by using smaller amounts of catalyst, operating at higher temperatures or different pressures, at greater residence times, or at different gas compositions.

Whereas changing catalyst volume or residence time rarely yields complications, changing temperature or pressure could introduce sintering. The properties of the catalyst should be measured both before and after deactivation and inlet and outlet streams should be analyzed by chromatography (qv) or spectrometry.

Around 1972, it was reasoned that the problem of catalyst deactivation could not always be entirely eliminated, but that continuous replacement of a portion of the catalyst bed during normal operation would allow continuing operation at high efficiency even in the presence of poisoning agents. Hence the fluidized bed was born (8). In some applications, fluidized-bed oxidation processes overcame poisoning, masking, and thermal aging. A process in which performance depends on the continuous attrition of the external surface of the catalyst particles, however, has many unattractive features, including the effort required for trapping, collecting, and disposing of the fine particulate released from the reactor (32).

At least one printer using catalytic oxidation has experienced relatively rapid catalyst deactivation, requiring replacement after useful lives as short as 3–6 months, when producing high quality printed matter using some of the most desirable lithographic printing plates and materials (34). It was observed that the use of phosphorus additives caused rapid deactivation of the conventional catalyst used to destroy the hydrocarbons in the solvent-laden air (SLA) discharged from the press dryer. A precious-metal catalyst containing platinum and palladium was being used in this application. It had replaced an earlier base metal catalyst, which showed rapid deactivation as a result of sulfur in the SLA, presumably introduced in the fuel used to fire the heatset dryer. It was found that the P concentration in the SLA might be as high as 0.16 ppm. The phosphorus concentration on the deactivated catalyst was found to be $1.4^{\circ}\%$ of the catalyst. The printer was urged by the supplier to find and eliminate the cause of the phosphorus contamination of the waste gas entering the catalyst bed. However, after it was determined that use of phosphorus-containing additives was crucial to many of the high quality printing jobs, attention was directed at the catalyst bed itself.

A catalyst with a substantially improved resistance to poisoning by phosphorus in catalytic oxidation applications was developed. In part, the catalyst in this program permitted printers to use photolithographic technology without paying an unreasonable cost in terms of frequent replacement of oxidation catalysts.

Catalyst Reactivation. Some catalytic systems are reported to have operated continuously for more than 10 years with little or no loss in control efficiency (5). In most processes catalysts inevitably lose activity, and when the activity has declined to a critical level, the catalyst needs to be discarded or regenerated (24). Regeneration is only possible when the deactivation is reversible by chemical washing or heat treatment or oxidation (20,24).

Thermal Treatment. A thermal treatment for catalyst regeneration is usually effective when deactivation is a result of coking or masking of the catalyst surface. Thermal treatment can usually be done on-site, by elevating the temperature of the catalyst bed by 50 to 100°C above the normal operating point and running at this oxidizing condition for a specified limited period of time (20). The elevated temperature vaporizes or oxidizes the organic compounds or char that may be masking the catalyst surface.

Physical Treatment. If inspection of the catalyst indicates deposits again, or if an excessive pressure drop across the catalyst is noted, then the catalytic bed may be lanced, on-site, using compressed air or water until the deposits are removed. Abrasion by contact with excessively high pressure from the compressed air should be avoided (20). If this treatment is combined with heating, hot spots or overtemperatures that could further deactivate the catalyst should be avoided (24). In many cases, periodic maintenance, removing the catalyst bed and blowing or washing off residues, has restored catalyst to original or near-original activity levels (5).

Chemical Treatment. The most involved regeneration technique is chemical treatment (20) which often follows thermal or physical treatment, after the char and particulate matter has been removed. Acid solution soaks, glacial acetic acid, and oxalic acid are often used. The bed is then rinsed with water, lanced with air, and dried in air. More involved is use of an alkaline solution such as potassium hydroxide, or the combination of acid washes and alkaline washes. The most complex treatment is a combination of water, alkaline, and acid washes followed by air lancing and drying. The catalyst should not be appreciably degraded by the particular chemical treatment used.

Analyses of a catalyst used in a process involving cleaning products and pigments and achieving a hydrocarbon destruction capacity of only $13^{\circ}\%$ showed deposition of P, Sn, Pb, and Na contaminants (20). Initial acid treatment increased the hydrocarbon destruction capacity from 13 to $63^{\circ}\%$. Alkaline treatment increased the capacity to $90^{\circ}\%$ of that new.

Exhaust Control Technologies

In addition to VOCs, specific industrial exhaust control technologies are available for nitrogen oxides, NO_x , carbon monoxide, CO , halogenated hydrocarbon, and sulfur and sulfur oxides, SO_x .

Nitrogen Oxides. Annual releases of nitrogen oxides (NO_x) into the atmosphere amounted to ca 550×10^6 t in the early 1990s. A number of states, in addition to California, regulate NO_x emissions (35). The production of nitrogen oxides can be controlled to some degree by reducing formation in the combustion system. The rate of NO_x formation for any given fuel and combustor design are controlled by the local oxygen concentration, temperature, and time history of the combustion products. Techniques employed to reduce NO_x formation are collectively referred to as combustion controls and U.S. power plants have shown that furnace modifications can be a cost-effective approach to reducing NO_x emissions. Combustion control technologies include operational modifications, such as low excess air, biased firing, and burners-out-of-service, which can achieve 20–30% NO_x reduction; and equipment modifications such as low NO_x burners, overfire air, and reburning, which can achieve 40–60% reduction (36). As of this writing,

approximately 600 boilers having 10,000 MW of capacity use combustion modifications to comply with the New Source Performance Standards (NSPS) for NO_x emissions (37).

When NO_x destruction efficiencies approaching 90% are required, some form of post-combustion technology applied downstream of the combustion zone is needed to reduce the NO_x formed during the combustion process. Three post-combustion NO_x control technologies are utilized: selective catalytic reduction (SCR); nonselective catalytic reduction (NSCR); and selective noncatalytic reduction (SNCR).

Selective Catalytic Reduction. Selective catalytic reduction (SCR) is widely used in Japan and Europe to control NO_x emissions (1). SCR converts the NO_x in an oxygen-containing exhaust stream to molecular N₂ and H₂O using ammonia as the reducing agent in the presence of a catalyst. NO_x removals of 90% are achievable. The primary variable is temperature, which depends on catalyst type (38). The principal components of an SCR system include the catalyst, the SCR reactor, the ammonia injection grid (AIG), the ammonia–air dilution system, the ammonia storage–vaporization system, the ammonia addition control system, and a continuous emissions monitoring system (39).

The AIG is used to uniformly inject diluted ammonia into the reactor. Uniform mixing of the ammonia into the flue gas is necessary to maintain catalyst performance at its highest level and to minimize ammonia leakage (ammonia slip) past the catalyst.

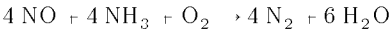
The ammonia–air dilution system dilutes the vaporized ammonia by a factor of 20 to 25 with air for better admixing through the AIG and to prevent explosive ammonia–air mixtures. Once the catalyst volume is selected, the NO_x removal is set by the NH₃/NO_x mole ratio at the inlet of the SCR system (39).

SCR was first developed in the United States in the late 1950s, targeted at nitric acid tail-gas exhausts, using precious-metal catalysts. In the mid-1970s, SCR entered widespread commercial use in Japan using base metal catalysts. The first SCR systems in Germany started up in 1986 (39), and German utilities are committed to installing SCR systems on the majority of oil- and coal-fired boilers to achieve between 60 and 90% NO_x reductions (39). By the end of 1992, there were about 120 SCR plants in service in Germany alone (40). A limited amount of experience has been documented in the United States (41,42), although commercial service began in 1985 and is expected to increase (43).

Performance criteria for SCR are analogous to those for other catalytic oxidation systems: NO_x conversion, pressure drop, catalyst system life, cost, and minimum SO₂ oxidations to SO₃. An optimum SCR catalyst is one that meets both the pressure drop and NO_x conversion targets with the minimum catalyst volume. Because of the interrelationship between cell density, pressure drop, and catalyst volume, a wide range of optional catalyst cell densities are needed for optimizing SCR system performance.

Reactions. The SCR process is termed selective because the ammonia reacts selectively with NO_x at temperatures >232°C in the presence of excess oxygen (44). The optimum temperature range for the SCR catalyst is determined by balancing the needs of the redox reactions.

SCR reactions

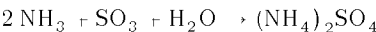
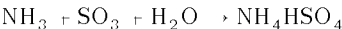


The NO reduction is the most important because NO₂ accounts for only 5–10% of the NO_x in most exhaust gases.

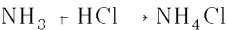
Ammonia oxidation reactions



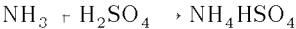
When sulfur dioxide is also present there are important side reactions in which SO₂ is oxidized to SO₃. The main side reaction in the SCR catalyst is the conversion of SO₂ to SO₃, thus facilitating the reaction above. The SO₃ in turn reacts with ammonia to form ammonium sulfates.



The formation of ammonium bisulfate is strongly temperature dependent. Formation is favored at the lower temperatures. The temperature at which ammonium bisulfate is not formed depends strongly on the SO₃ concentration in the exhaust gas. The temperature needed to minimize bisulfate formation has been reported to increase by about 15°C (around about 350°C) when the SO₃ concentration increases from 5 to 15 ppm (23). The formation of the bisulfate is reversible, ie, if the temperature is raised to 20°C above the minimum temperature, the reaction is shifted to result in the decomposition of the bisulfate formed. When chlorides are present, ammonium chlorides can be formed:



When sulfuric acid is present, ammonium bisulfate can be formed:



These various reactions should be minimized to avoid plugging the catalyst and to prevent fouling of the downstream air preheaters, when these components condense from the gas at the lower temperatures.

The SCR Process. The first step in the SCR reaction is the adsorption of the ammonia on the catalyst. SCR catalysts can adsorb considerable amounts of ammonia (45). However, the adsorption must be selective and high enough to yield reasonable cycle times for typical industrial catalyst loadings, ie, uptakes in excess of 0.1% by weight. The rate of adsorption must be comparable to the rate of reaction to ensure that suitable fronts are formed. The rate of desorption must be slow. Ideally the adsorption isotherm is rectangular. For optimum performance, the reaction must be irreversible and free of side reactions.

It has been suggested that the first step, weak coadsorption of NO and O₂ on a reduced vanadium site, may represent the slow step in the mechanism. Subsequent formation of a N₂O₃-like intermediate could be a fast step, because it is known that in the gas phase the equilibrium of NO–NO₂ and N₂O₃ is established within microseconds (35).

At low temperatures the SCR reactions dominate and nitrogen oxide conversion increases with increasing temperature. But as temperature increases, the ammonia oxidation reactions become relatively more important. As the temperature increases further, the destruction of ammonia and generation of nitrogen oxides by the oxidation reactions causes overall nitrogen oxide conversion to reach a plateau then decreases with increasing temperatures. Examples are shown in Figure 7 (44).

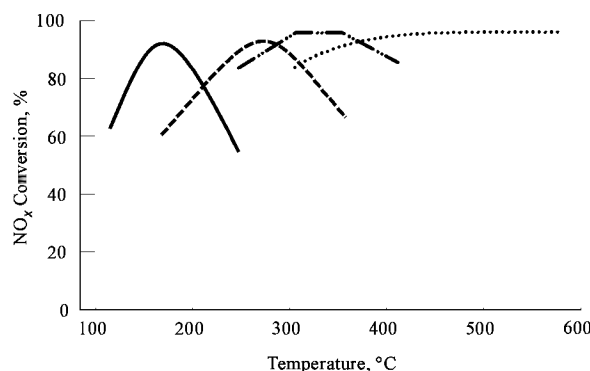


Fig. 7. NO_x reduction vs temperature for SCR catalysts: (—) CAT LT-1, (---) CAT LT-2, (- · -) CAT VNX, and (·····) CAT 2NX (41).

In the SCR process, ammonia, usually diluted with air or steam, is injected through a grid system into the flue exhaust stream upstream of a catalyst bed (37). The effectiveness of the SCR process is also dependent on the NH₃ to NO_x ratio. The ammonia injection rate and distribution must be controlled to yield an approximately 1:1 molar ratio. At a given temperature and space velocity, as the molar ratio increases to approximately 1:1, the NO_x reduction increases. At operations above 1:1, however, the amount of ammonia passing through the system increases (38). This ammonia slip can be caused by catalyst deterioration, by poor velocity distribution, or inhomogeneous ammonia distribution in the bed.

Types of SCR Catalysts. The catalysts used in the SCR were initially formed into spherical shapes that were placed either in fixed-bed reactors for clean gas applications or moving-bed reactors where dust was present. The moving-bed reactors added complexity to the design and in some applications resulted in unacceptable catalyst abrasion. As of 1993 most SCR catalysts are either supported on a ceramic or metallic honeycomb or are directly extruded as a honeycomb (1). A typical honeycomb block has face dimensions of 150 by 150 mm and can be as long as one meter. The number of cells per block varies from 20 by 20 up to 45 by 45 (39).

No SCR catalyst can operate economically over the whole temperature range possible for combustion systems. As a result, three general classes of catalysts have evolved for commercial SCR systems (44): precious-metal catalysts for operation at low temperatures, base metals for operation at medium temperatures, and zeolites for operation at higher temperatures.

The precious-metal platinum catalysts were primarily developed in the 1960s for operation at temperatures between about 200 and 300°C (1,38,44). However, because of sensitivity to poisons, these catalysts are unsuitable for many combustion applications. Variations in sulfur levels of as little as 0.4 ppm can shift the catalyst required temperature window completely out of a system's operating temperature range (44). Additionally, operation with liquid fuels is further complicated by the potential for deposition of ammonium sulfate salts within the pores of the catalyst (44). These low temperature catalysts exhibit NO_x conversion that rises with increasing temperature, then rapidly drops off, as oxidation of ammonia to nitrogen oxides begins to dominate the reaction (see Fig. 7).

The most popular SCR catalyst formulations are those that were developed in Japan in the late 1970s comprised of base metal oxides such as vanadium pentoxide [1314-62-1], V₂O₅, supported on titanium dioxide [13463-67-7], TiO₂ (1). As for low temperature catalysts, NO_x conversion rises with increasing temperatures to a plateau and then falls as ammonia oxidation begins to dominate the SCR reaction. However, peak conversion occurs in the temperature range between 300 and 450°C, and the fall-off in NO_x conversion is more gradual than for low temperature catalysis (44).

A family of zeolite catalysts has been developed, and is being increasingly used in the United States in SCR applications. Zeolites which can function at higher temperatures than the conventional catalysts, are claimed to be effective over the range of 300 to 600°C, having an optimum temperature range from 360 to 580°C (37,38). However, ammonia oxidation to NO_x begins around 450°C and is predominant at temperatures in excess of 500°C. Zeolites suffer the same performance and potential damage problems as conventional catalysts when used outside the optimum temperature range. In particular, at around 550°C the zeolite structure may be irreversibly degraded because of loss of pore density. Zeolite catalysts have not been continuously operated commercially at temperatures above 500°C (37).

Using zeolite catalysts, the NO_x reduction takes place inside a molecular sieve ceramic body rather than on the surface of a metallic catalyst (see MOLECULAR SIEVES). This difference is reported to reduce the effect of particulates, soot, SO₂, SO₃ conversions, heavy metals, etc, which poison, plug, and mask metal catalysts. Zeolites have been in use in Europe since the mid-1980s and there are approximately 100 installations on stream. Process applications range from use of natural gas to coal as fuel. Typically, nitrogen oxide levels are reduced 80 to 90% (37).

Catalyst Selection. For an SCR application, catalyst selection depends largely on the temperature of the flue gas being treated. A given catalyst exhibits optimum performance within a temperature range of about 30 to 50°C. Below this optimum temperature range, the catalyst activity is greatly reduced, allowing unreacted ammonia to slip through. Above this range, ammonia begins to be oxidized to form additional NO_x. Operations having adequate temperature controls are important, as are uniform flue gas temperatures (37,38).

Problems. A number of difficulties in utilizing SCR operations have been identified. Problems in European installations are of particular interest because SCR systems are subjected to conditions not experienced in Japan, but also encountered in the United States. Difficulties include matching the NH₃ injection pattern to the nonuniform flow of NO_x in the ductwork ahead of the SCR reactor; inability to optimize the NH₃ injection rate by feedback control of slip ammonia for lack of a reliable NH₃ monitor; erosion and plugging on units retrofitted to boilers that fire high ash coal; catalyst deactivation caused by arsenic poisoning on wet-bottom units that recycle flyash; process control under load swings; and spent catalyst disposal. For medium to high sulfur coals, the potential exists for accelerated catalyst deactivation caused by sulfur poisoning and contamination by trace metals in flyash, and deposit buildup and corrosion of the air heater.

For boilers, SCR DENO_x plants can be installed at the exhaust exit just before the air preheater. These are called high dust plants (17 g dust per cubic meter) (23) because the flue gas still contains volatile trace elements as well as flyash particles. When the SCR system is installed after the flue gas desulfurization (FGD) system, it is called a low dust or tail-end plant. The high dust plant has the advantages of not requiring any additional energy because of the high temperatures present. The low dust plant requires a regenerative heat exchanger, but requires less catalyst because (40): there is less dust to contaminate the bed, potential catalyst poisons are removed from the flue gas by the FGD system, and a high activity catalyst can be used when only low concentrations of poisons such as SO₂ remain after the flue gas system.

Other problems that can be associated with the high dust plant can include alkali deterioration from sodium or potassium in the stack gas deposition on the bed, calcium deposition, when calcium in the flue gas reacts with sulfur trioxide, or formation and deposition of ammonium bisulfate. In addition, plugging of the air preheater as well as contamination of flyash and FGD wastewater discharges by ammonia are avoided if the SCR system is located after the FGD (23).

A significant problem area for initial SCR systems has been the continuous emission monitoring (CEM) systems. In power plants, all sites equipped with CEM systems report the highest failure frequency. The CEM systems are the most labor intensive component, requiring as much as full-time attention from one technician. At one power plant CEM systems were responsible for 100% of 73 reported SCR system shutdowns (38). As CEM systems improve, these concerns may disappear.

Nonselective Catalytic Reduction. Hydrocarbons, hydrogen, or carbon monoxide can be used as reducing agents for NO_x in applications where the exhaust oxygen concentration is low, as it is in fuel rich-burn reciprocating engines, where it is less than 1% o, and in nitric acid plants, when it is from 2 to 3% o. This approach is called nonselective catalytic reduction (NSCR). In some applications, the oxygen must be removed from the feed stream prior to the catalyst (35). An oxygen sensor in the exhaust stream signals the air–fuel delivery system to adjust the air–fuel ratio so it is just slightly fuel-rich, having enough reducing agent present to react with all the oxygen and nitrogen oxides (1).

Nonselective catalytic reduction systems are often referred to as three-way conversions. These systems reduce NO_x, unburned hydrocarbon, and CO simultaneously. In the presence of the catalyst, the NO_x are reduced by the CO resulting in N₂ and CO₂ (37). A mixture of platinum and rhodium has been generally used to promote this reaction (37). It has also been reported that a catalyst using palladium has been used in this application (1). The catalyst operation temperature limits are 350 to 800°C, and 425 to 650°C are the most desirable. Temperatures above 800°C result in catalyst sintering (37). Automotive exhaust control systems are generally NSCR systems, often shortened to NCR.

Typically NO_x conversion ranges from 80 to 95% o and there are corresponding decreases in CO and hydrocarbon concentrations. Potential problems associated with NSCR applications include catalyst poisoning by oil additives, such as phosphorus and zinc, and inadequate control systems (37).

Carbon Monoxide. Carbon monoxide is emitted by gas turbine power plants, reciprocating engines, and coal-fired boilers and heaters. CO can be controlled by a precious-metal oxidation catalyst on a ceramic or metal honeycomb. The catalyst promotes reaction of the gas with oxygen to form CO₂ at efficiencies that can exceed 95% o. CO oxidation catalyst technology is broadening to applications requiring better catalyst durability, such as the combustion of heavy oil, coal (qv), municipal solid waste (qv), and wood (qv). Research is underway to help cope with particulates and contaminants, such as flyash and lubricating oil, in gases generated by these fuels (1).

CO conversion is a function of both temperature and catalyst volume, and increases rapidly beginning at just under 100°C until it reaches a plateau at about 150°C. But, unlike NO_x catalysts, above 150°C there is little benefit to further increasing the temperature (44). Above 150°C, the CO conversion is controlled by the bulk phase gas mass transfer of CO to the honeycomb surface. That is, the catalyst is highly active, and its intrinsic CO removal rate is exceedingly greater than the actual gas transport rate (21). When the activity falls to such an extent that the conversion is no longer controlled by gas mass transfer, a decline of CO conversion occurs, and a suitable regeneration technique is needed (21).

It has been reported that below about 370°C, sulfur oxides reversibly inhibit CO conversion activity. This inhibition is greater at lower temperatures. CO conversion activity returns to normal shortly after removal of the sulfur from the exhaust (44). Above about 315°C, sulfur oxides react with the high surface area oxides to disperse the precious-metal catalytic agents and irreversibly poison CO conversion activity.

Catalyst contamination from sources such as turbine lubricant and boiler feed water additives is usually much more severe than deactivation by sulfur compounds in the turbine exhaust. Catalyst formulation can be adjusted to improve poison tolerance, but no catalyst is immune to a contaminant that coats its surface and prevents access of CO to the active sites. Between 1986 and 1990 over 25 commercial CO oxidation catalyst systems operated on gas turbine cogeneration systems, meeting both CO conversion (40 to 90% o) and pressure drop requirements.

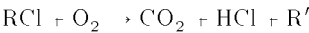
Halogenated Hydrocarbons. Destruction of halogenated hydrocarbons presents unique challenges to a catalytic oxidation system (45–51). The first step in any control strategy for halogenated hydrocarbons is recovery and recycling (45). However, even upon full implementation of economic recovery steps, significant halocarbon emissions can remain. In other cases, halogenated hydrocarbons are present as impurities in exhaust streams (45). Impurity sources are often intermittent and dispersed.

The principal advantage of a catalytic oxidation system for halogenated hydrocarbons is in operating cost savings. Catalytically stabilized combustors improve the incineration conditions, but still must employ very high temperatures as compared to VOC combustors; eg, carbon tetrachloride [56-23-5], CCl₄, has a 40-fold lower heat of combustion than a typical organic vapor such as toluene [108-88-3], thus CCl₄ requires much more supplemental fuel to burn than do typical organics (45). Alternatively, the low temperature catalytic oxidation process is typically designed for a maximum adiabatic temperature rise of only 200°C. This would correspond to only about 1500 ppm of an organic compound in the exhaust stream. But, with the lower heat of combustion, up to 40,000 ppm of carbon tetrachloride could be treated in the same temperature rise, or with less dilution air.

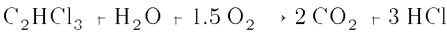
By-Product Formation. The presence of halogenated hydrocarbons dramatically increases the yield of aldehydes from the oxidation process (45). For example, in the partial oxidation of methane on a PdO sponge catalyst (19), methylene chloride, CH₂Cl₂, was added in pulses to the inlet gas, which also contained the methane. Methane oxidation was strongly inhibited and formaldehyde was formed. The formaldehyde production continued even after the methylene chloride addition was stopped, suggesting a strong interaction of chlorine with the catalyst. However, pulses of pure CH₄ plus oxygen gradually restored the original activity to the catalyst, indicating that the effect of this interaction was reversible.

Catalyst Deactivation. Catalyst deactivation (45) by halogen degradation is a very difficult problem particularly for platinum (PGM) catalysts, which make up about 75% o of the catalysts used for VOC destruction (10). The problem may well lie with the catalyst carrier or washcoat. Alumina, for example, a common washcoat, can react with a chlorinated hydrocarbon in a gas stream to form aluminum chloride which can then interact with the metal. Fluid-bed reactors have been used to offset catalyst deactivation but these are large and costly (45).

Catalytic Reaction. The desired reaction of the chlorine group on a chlorinated hydrocarbon is



It is important to produce HCl rather than elemental chlorine, Cl₂, because HCl can be easily scrubbed out of the exhaust stream, whereas Cl₂ is very difficult to scrub from the reactor off-gas. If the halogenated hydrocarbon is deficient in hydrogen relative to that needed to produce HCl, low levels of water vapor may be needed in the entering stream (45) and an optional water injector may be utilized. For example, trichloroethylene [79-01-6], C₂HCl₃, and carbon tetrachloride require some water vapor as a source of hydrogen (45).



Groundwater contaminated with chlorinated hydrocarbons is being remediated by a conventional air stripper or a rotary stripper, producing an air stream containing the halogenated hydrocarbon vapors and saturated with water vapor (45), which is then passed through a catalyst bed.

At least two catalytic processes have been used to purify halogenated streams. Both utilize fluidized beds of probably nonnoble metal catalyst particles. One has been estimated to oxidize >9000 t yr of chlorinated wastes from a vinyl chloride monomer plant (45). Several companies have commercialized catalysts which are reported to resist deactivation from a wider range of halogens. These newer catalysts may allow the required operating temperatures to be reduced, and still convert over 95% o of the halocarbon, such as trichlorethylene, from an exhaust stream. Conversions of C-1 chlorocarbons utilizing an Englehardt HDC catalyst are shown in Figure 8. For this system, as the number of chlorine atoms increases, the temperatures required for destruction decreases.

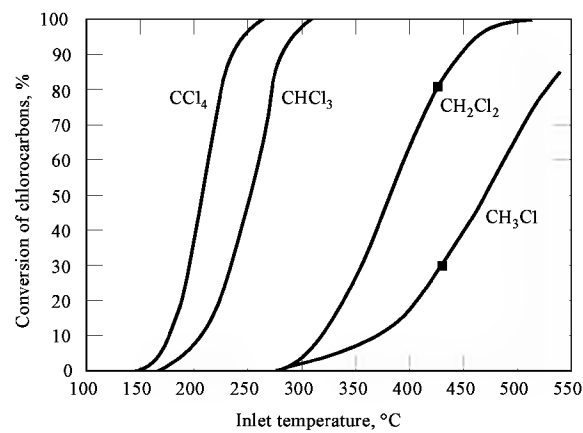


Fig. 8. Destruction of C-1 chlorocarbons over HDC in the presence of 1.5% H₂O in air at 15,000 h⁻¹ GHSV at STP. Chlorocarbon concentrations are CCl₄, 900 ppmv; CHCl₃, 500 ppmv; CH₂Cl₂, 800 ppmv; and CH₃Cl, 600 ppmv.

Uses

Catalytic oxidation of exhaust streams is increasingly used in those industries involved in the following (13,15,17):

- Surface coatings*
 - aerospace
 - automobile
 - auto refinishing
 - can coating
 - coil coating
 - fabric coating
 - large appliances
 - marine vessels
 - metal furniture
 - paper coating
 - plastic parts coating
 - wire coating and enameling
 - wood furniture
- Printing inks*
 - flexographic
 - lithographic
 - rotogravure
 - screen printing
- Solvent usage*
 - adhesives
 - disk manufacture
 - dry cleaning
 - fiber glass manufacture
 - food tobacco manufacture
 - metal cleaning
 - pharmaceutical
 - photo finishing labs
 - semiconductor manufacture
- Chemical and petroleum processes*
 - cumene manufacture
 - ethylene oxide manufacture
 - acrylonitrile manufacture
 - caprolactam manufacture
 - maleic anhydride manufacture
 - monomer venting
 - phthalic anhydride manufacture
 - paint and ink manufacture
 - petroleum product refining
 - petroleum marketing
 - resin manufacture
 - textile processing
- Industrial/ commercial processes*
 - aircraft manufacture
 - aseptic packaging
 - asphalt blowing
 - automotive parts manufacture
 - breweries wineries
 - carbon fiber manufacture
 - catalyst regeneration
 - coffee roasting
 - commercial charbroiling
 - electronics manufacture
 - film coating

- filter paper processing
- food deep frying
- gas purification
- glove manufacture
- hospital sterilizers
- peanut and coffee roasting
- plywood manufacture
- rubber processing
- spray painting
- tire manufacture
- wood treating
- Engines*
- diesel engines
- lean burn internal combustion
- natural gas compressors
- oil field steam generation
- rich burn internal combustion
- gas turbine power generation
- Cross media transfer*
- air stripping
- groundwater cleanup
- soil remediation (landfills)
- hazardous waste treatment
- odor removal from sewage gases

The most important factors affecting performance are operating temperature, surface velocity, contaminant concentration and composition, catalyst properties, and the presence or absence of poisons or inhibitors.

Air Stripping of Groundwater. Treatment of exhaust streams from the air stripping of contaminated groundwater is a particular challenge, because the emissions from air stripping units may consist of a complex mixture of both fuel and solvent fractions (6). The catalytic oxidation of any given compound is generally negatively impacted by the presence of others in mixtures, and higher catalyst bed operating temperatures are required to achieve adequate destruction.

Some catalysts exposed to air stripping off-gas were subject to deactivation. However, using a catalytic oxidizer at a U.S. Coast Guard facility (Traverse City, Mich.) for the destruction of benzene, toluene, and xylene stripped from the groundwater, the catalytic oxidization unit operated at 260 to 315°C, and was able to achieve 90% destruction efficiency (see GROUNDWATER MONITORING).

Printing and Graphic Arts. In the graphic arts industry, the catalyst in the oxidizer needs to be monitored regularly because it is susceptible to contamination by phosphorus from fountain solutions, silica from silicone gloss enhancer sprays, and chlorides from chlorinated solvents or blanket wash solutions. Phosphorus and silica accumulate most rapidly on the leading edge of the catalyst bed, deactivating the catalyst by masking the precious metals. In a fluidized-bed configuration, the catalyst surface is continually renewed by abrasion and the problem of masking the catalyst surface with silicones is avoided.

Chemical Processing.

Terephthalic Acid Production. The control of exhaust from production of pure terephthalic acid (PTA) has been a challenge (see CARBOXYLIC ACIDS) (52). Eight million metric tons of PTA are produced annually worldwide for use primarily in high grade polyester fiber production. Based on a *para*-xylene feedstock, vent gases from the process contain such by-products as methyl acetate, organic acids, and often methyl bromide. These exhausts have been estimated to total 34,000 m³ worldwide. Historically, the presence of the methyl bromide limited the use of fixed-bed catalytic oxidation as a control technology using precious-metal catalysts. Thus base metal catalysts in fluidized-bed reactors have been the primary catalytic technology of choice. In this application, the continuous abrasion of the outer layer of the catalyst particles exposes a fresh surface of unpoisoned material to the reactants, allowing the catalyst to effectively treat the exhaust stream.

In the late 1980s, however, the discovery of a noble metal catalyst that could tolerate and destroy halogenated hydrocarbons such as methyl bromide in a fixed-bed system was reported (52,53). The products of the reaction were water, carbon dioxide, hydrogen bromide, and bromine. Generally, a scrubber would be needed to prevent downstream equipment corrosion. However, if the focus of the control is the VOCs and the CO rather than the methyl bromide, a modified catalyst formulation can be used that is able to tolerate the methyl bromide, but not destroy it. In this case the methyl bromide passes through the bed unaffected, and designing the system to avoid downstream effects is not necessary. Destruction efficiencies of hydrocarbons and CO of better than 95% have been reported, and methyl bromide destructions between 0 and 85% (52).

Latex Monomer Production. ARI Technologies, Inc. has introduced a catalyst system which, it is claimed, can operate at an average bed temperature of 370°C while achieving conversion efficiency in excess of 99.99% on exhaust streams from latex monomer production (see LATEX TECHNOLOGY).

Acrylonitrile Manufacture. In the manufacture of acrylonitrile (qv), off-gases containing from 1–3% of CO plus various hydrocarbons are emitted. Catalytic beds of platinum-group metals are used to reduce the regulated compounds to acceptable levels. Close attention to bed design is required to prevent the formation of appreciable quantities of NO_x caused by the fixation of combustion-air nitrogen. Some NO_x also is produced from fuel nitrogen by oxidation. Because of the high thermal energy content of the off-gases, considerable heat recovery is possible in abating acrylonitrile plant emissions (see HEAT EXCHANGE TECHNOLOGY).

Vinyl Monomer Manufacturing. Process vent gases containing small quantities of halogenated hydrocarbons and substantial quantities of nonhalogenated hydrocarbons have been successfully reduced to comply with regulatory objectives in large-scale laboratory–pilot catalytic fume abaters having satisfactory long-term catalyst performance. The design freedoms offered by precious metals on ceramic honeycomb support catalysts have been demonstrated in equipment that utilizes the heat energy resulting from the substantial exotherm of the nonhalogenated hydrocarbon oxidation to preheat the exhaust gases. Fuel consumption is thereby minimized.

Coatings Industries. Surface coating processes (qv) produce similar air pollution problems in a number of different industries (see also COATINGS).

Can Manufacturing. An internal coating is necessary to protect the purity and flavor of can contents for beverages or any edible product that might react with the container metal (see BEER; CARBONATED BEVERAGES; FOOD PACKAGING). Both the exterior decorative and interior sanitary coatings are applied to the metal surface by rolls or spray guns using a solvent vehicle. Catalytic oxidation systems are used by the principal can manufacturers to treat coatings exhaust streams. The can manufacturers' industry is estimated to utilize more catalysts than any of the other surface coating industries.

A large number of diverse solvents are used in exterior and interior coatings in plants for manufacturing both three- and two-piece cans. Most of the organic solvents are found in the cure-oven exhausts at concentrations of 2–16% of the lower explosive limit (LEL). The oven exhaust volumes are usually 1–35 m³ s. When burned, these concentrations of combustibles provide an exotherm of 30 to 220°C. The heat that is released is used for preheating the incoming effluent and or heating the cure oven by recycling the hot, cleaned gases to the supply blowers or by heating makeup air by heat exchange. A few plants use the heat of the cleaned exhaust to produce hot water for the two-piece can line washers, hot air for dry-off ovens, or building space heating. For example, one large can company utilizes the heat energy contained in the stream, leaving some of their catalytic fume abaters to supply all the heat energy required by the oven's heating zones, which have no burners. The fuel energy supplied to the catalytic fume abater is less than would be needed to heat the oven if the solvent fumes were exhausted directly to the atmosphere without use of the fume abater. The exhaust rate of the oven is adjusted to maintain a

solvent concentration of at least 8° of the LEL, equivalent to a 110°C temperature differential (see ENERGY MANAGEMENT).

The various reaction rate properties of the different solvents influence the design of a catalytic reactor. For example, for a specific catalyst bed design, an effluent stream containing a preponderance of monohydric alcohols, aromatic hydrocarbons, or propylene requires a lower catalyst operating temperature than that required for solvents such as isophorone and short-chain acetates.

Design considerations and costs of the catalyst, hardware, and a fume control system are directly proportional to the oven exhaust volume. The size of the catalyst bed often ranges from 1.0 m³ at 0°C and 101 kPa per 1000 m³ min of exhaust, to 2 m³ for 1000 m³ min of exhaust. Catalyst performance at a number of can plant installations has been enhanced by proper maintenance. Annual analytical measurements show reduction of solvent hydrocarbons to be in excess of 90° for 3–6 years, the equivalent of 12,000 to 30,000 operating hours. When propane was the only available fuel, the catalyst cost was recovered by fuel savings (vs thermal incineration prior to the catalyst retrofit) in two to three months. In numerous cases the fuel savings paid for the catalyst in 6 to 12 months.

Can manufacturers often regenerate the catalyst beds on an annual or biannual basis during a weekend downtime. Both air lancing and an aqueous bath are utilized to remove noncombustible particulates that mask the active sites. Frequently, condensed organic material on the catalyst is removed by short-term (4–6 h) heating excursions to 370 or 430°C; the organic matter is removed much like a self-cleaning oven. The gaseous and organic smoke, which is usually evolved from the first few cm of catalyst bed depth, is oxidized in the latter part of the bed. If allowed to operate too long at temperatures that promote condensation, high boiling organic compounds, the subsequent carbon char that is formed, may require temperatures of 480 to 540°C to convert the carbon to carbon monoxide for subsequent oxidation. The higher temperatures required for burn-offs should be approached in small (0–30°C) increments to bring about slow evolution and partial oxidation; this prevents autogenous combustion of local high concentrations of combustible material.

Day-to-day operating techniques that are employed by one large can manufacturer and are intended to prevent organic condensation are dictated by the use of a low cost, well-established, sanitary coating for beer and beverage three-piece cans. Polybutadiene and other sanitary coatings may have volatile resin monomers entrained in the oven atmosphere as a result of rapid evaporation of solvent before polymerization takes place. A short (4 to 6 h) heating excursion up to a catalyst inlet temperature of 400°C after use of the coating usually burns off any condensed organic materials. It has become standard practice in some plants to turn the catalytic afterburner up to 370°C for these coatings vs the normal 315°C operating temperature for vinyls, acrylics, etc.

Coil Coating. Coil coating is the prefinishing of many sheet metal items with protective and decorative coatings that are applied by roll coating on one or both sides of a fast-moving metal strip. The metal strip (from 13 mm to 1.7 m in width) unwinding from a coil travels at rates of 30–150 m min through the coating applicator rolls and bake ovens. It is rewound into a coil for transport to a forming operation for products that are to be used in cans, appliances, industrial and residential siding, shelving, cars, gutters, downspouts, etc. The source of hydrocarbon emissions in coil coating is the coating application area and the cure-oven exhaust. The coatings include primers, finishes, and metal protective (5 µm) films or backers.

The increasing use of siliconized coatings for weather durability caused severe masking problems for the all-metal, filter mesh-like catalyst elements available in the 1970s. Interest in catalytic afterburners increased when dispersed-phase precious metal–alumina-on-ceramic honeycomb catalysts offered economically attractive results.

The hot (260–370°C) oven exhaust can be oxidized catalytically without preheat, but when the coater-area exhaust (at room temperature) is combined with the oven exhaust, a preheat burner becomes necessary. The greatest energy savings potential having the least capital investment is obtained by recycling a portion of the hot, cleaned exhaust to the oven. This principle has been demonstrated at a number of can manufacturing plants and at least four coil coating facilities. One operator preheats oven make-up air which has been taken from the oven cooler section by means of a heat exchanger; whereas others recycle directly to the oven. In one case, the heat energy for the dry-off oven is supplied by the catalytic incinerator exhaust (482°C) remaining after supplying most of the heat energy to operate the four zones in the oven. The concentration of solvents in the exhaust is about 12° LEL (167°C temperature differential). The net fuel energy consumption is about 20° of that required to fire the paint, bake, and dry-off ovens without fume control.

Coil coaters operate equipment continuously and, in most cases, operate catalytic fume abaters 6000–7000 h yr. Under these conditions the anticipated catalyst life is years, with an annual aqueous solution cleaning. However, the catalyst may last no more than two years if frequent maintenance is needed, such as in-place air lancing every 60 to 90 days to remove noncombustible particulates. Frequent maintenance may be needed if coatings such as siliconized polyester (15–40° silicones) comprise 30° of the coatings put through the system.

Filter Paper Processing. In the fabrication of fuel oil and air filters for vehicles such as motorcycles and diesel locomotives, heat processing of the filter paper is required to cure the resin (usually phenolic) with which the paper (qv) is impregnated (see PHENOLIC RESINS). The cure-oven exhaust, which contains water vapor, alcohols, and dimers and trimers of phenol, produces a typical blue haze aerosol having a pungent odor. The concentration of organic substances in the exhaust is usually rather low.

The paper-impregnation drying oven exhausts contain high concentrations (10–20° LEL) of alcohols and some resin monomer. Vinyl resins and melamine resins, which sometimes also contain organic phosphate fire retardants, may be used for air filters. The organic phosphates could shorten catalyst life depending on the mechanism of reduction of catalyst activity. Mild acid leaching removes iron and phosphorus from partially deactivated catalyst and has restored activity in at least one known case.

Catalysis is utilized in the majority of new paper filter cure ovens as part of the oven recirculation burner system which is designed to keep the oven interior free of condensed resins and provide an exhaust without opacity or odor. The application of catalytic fume control to the exhaust of paper-impregnation dryers permits a net fuel saving by oxidation of easy-to-burn methyl or isopropyl alcohol, or both, at adequate concentrations to achieve a 110–220°C exotherm.

BIBLIOGRAPHY

"Exhaust Control, Industrial" in *ECT* 3rd ed., Vol. 9, pp. 511–541, by J. H. Conklin, D. M. Sowards, and J. H. Kroehling, E. I. du Pont de Nemours & Co., Inc.

1. R. J. Farrauto, R. M. Heck, and B. K. Speronnelo, *C & EN*, 34–44 (Sept. 7, 1992).
2. C. C. Lewis, *CPI Purch.* 29–33 (Aug. 1991).
3. J. C. Summers, J. E. Sawyer, and A. C. Frost "The 1990 Clean Air Act and Catalytic Emission Control Technology for Stationary Sources," in R. G. Silver, J. E. Sawyer, and J. C. Summers, eds., *Catalytic Control of Air Pollution: Mobile and Stationary Sources*," ACS Symposium Series 495, 1992.
4. R. L. Berglund and C. T. Lawson, "Pollution Prevention in the Chemical Process Industries," *Chem. Eng.* (Sept. 1991).
5. *Catalytic Control of VOC Emissions, A Cost Effective Viable Technology for Industrial, Commercial and Waste Processing Facilities*, Manufacturers of Emission Controls Association, Washington, D.C., 1992.
6. M. Kosusko and C. M. Nunez, "Destruction of Volatile Organic Compounds Using Catalytic Oxidation," *JAWMA* **40**(2) (Feb. 1990).
7. D. R. van der Vaart, W. M. Vatvuk, and A. H. Wehe *JAWMA* **41**(1), 92–98 (Jan. 1991).
8. L. C. Hardison and E. J. Dowd, *Chem Eng Prog.* 31–35 (Aug. 1977).
9. R. J. Martin, R. E. Smyth, and J. T. Schofield, "Elimination of Petroleum Industry Air Toxic Emissions with a Flameless Thermal Oxidizer," *Petro-Safe '93*, Houston, Tex., Jan. 29, 1993.
10. G. Parkinson, *Chem Eng.* 37–43 (July, 1991).
11. A. F. Hodel, *Chem Proc.*, 88–90 (June 1992).
12. R. Yarrington and L. Morris, "The VOC-Incinerator Option," *Indust. Finish.* (Mar. 1992).
13. J. Chen, R. M. Heck, and R. J. Farraoto *Catal. Today* **11**, 517–545 (1992).
14. M. S. Jennings, N. E. Krohn, and R. S. Berry, *Control of Industrial VOC Emissions by Catalytic Incineration*, Vol. 1, U.S. Environmental Protection

Agency, Research Triangle Park, N.C., July 1984.

15. E. J. Dowd, W. M. Sheffer, and G. E. Addison, "A Historical Perspective on the Future of Catalytic Oxidation of VOCs," paper 92-109.03, *85th Annual Meeting of Air and Waste Management Association*, Kansas City, Mo., June 21–26, 1993.

16. D. M. VanBenschoten, "On-Site Pilot Testing Demonstrates Catalytic Emission Control Technology for New VOC Applications," paper 92-109.01, presented at *85th Annual AWWMA Meeting & Exhibition*, Kansas City, Mo., June 21–26, 1992.

17. J. J. Spivey, *Ind. Eng. Chem. Res.* **26**, 2165–2180 (1987).

18. K. R. Bruns, "Use of Catalysts for VOC Control," presented at the *New England Environmental Expo*, Boston, Mass., Apr. 10–12, 1990.

19. J. J. Spivey and J. B. Butt, *Catal. Today* **11**, 465–500 (1992).

20. R. M. Heck, M. Durilla, A. G. Bouney, and J. M. Chen "Air Pollution Control—Ten Years Operating Experience with Commercial Catalyst Regeneration", paper presented at the *81st APCA Annual Meeting & Exhibition*, Dallas, Tex., June 19, 1988.

21. R. M. Heck, J. M. Chen, and M. F. Collins "Oxidation Catalyst for Cogeneration Applications—Regeneration of Commercial Catalyst," paper 90-105.1, presented at the *83rd Annual AWWMA Meeting & Exhibition*, Pittsburgh, Pa., June 24–29, 1990.

22. R. E. Kenson, "Control of Volatile Organic Emissions," Bulletin 1015, Series 1000, Met-Pro Corp., 1981.

23. W. L. Prins and Z. L. Nuninga, *Catal. Today* **16**, 187–105 (1993).

24. C. H. Bartholomew, *Chem. Eng.*, 96–119 (Nov. 12, 1984).

25. A. C. Frost and co-workers, *Environ. Sci. Technol.* **25**(12), 2065–2070 (1991).

26. D. Cicciella and B. Holt *Environ. Protec.*, 41–47 (Sept. 1992).

27. K. J. Herbert, "Catalysts for Volatile Organic Control in the 1990's," presented at the *1990 Incineration Conference*, San Diego, Calif., May 14–18, 1990.

28. M. A. Palazzolo, J. I. Steinmetz, D. L. Lewis, and J. F. Beltz "Parametric Evaluation of VOC HAP Destruction via Catalytic Incineration," U.S. Environmental Protection Agency, Report no. EPA 600 S2-85 041, Research Triangle Park, N.C., July 1985.

29. S. K. Gangwal, M. E. Mulling, J. J. Spivey, and P. R. Caffrey, *Appl. Catal.* **36**, 231–247 (1988).

30. M. A. Palazzolo, and C. L. Jamgonhian "Destruction of Chlorinated Hydrocarbons by Catalytic oxidation," U.S. Environmental Protection Agency, Report EPA 600 82-86 079, Research Triangle Park, N.C., Jan. 1987.

31. Difford and Spensor, *Chem. Eng. Prog.* **71**, 31 (1975).

32. G. R. Lester, "Catalytic Destruction of Hazardous Halogenated Organic Chemicals," presented at the *82nd Annual AWWMA Meeting & Exhibition*, Anaheim, Calif., June 25–30, 1992.

33. S. E. Wanke and P. C. Flynn, *Cat. Rev. Sci. Eng.* **12**, 93 (1975).

34. G. R. Lester, and J. C. Summers "Poison-Resistant Catalyst for Purification of Web Offset Press Exhaust," presented at the *Air Pollution Control Association, 81st Annual Meeting*, Dallas, Tex., June 19–24, 1988.

35. F. Luck and J. Roiron, *Catal. Today* **4**, 205–218 (1989).

36. E. Cichanowicz, *Power Eng.*, 36–38 (Aug. 1988).

37. L. M. Campbell, D. K. Stone, and G. S. Shareef, *Sourcebook: NO_x Control Technology Data*, EPA Report NO. EPA600 S2-91 029., Washington, D.C., Aug. 1991.

38. G. S. Shareef, D. K. Stone, K. R. Ferry, K. L. Johnson, and K. S. Locke "Selective Catalytic Reduction NO_x Control for Small Natural Gas-Fired Prime Movers," paper 92-136.06, in Ref. 16.

39. J. R. Donnelly and B. Brown "Joy Kawasaki Selective Reduction DE-NOX Technology," paper No 89-96B.6, in Ref. 32.

40. H. Gutberlet and B. Schallert, *Catal. Today* **16**, 207–236 (1993).

41. R. D. Walloch, *Oil Gas J.*, 39–41 (June 13, 1988).

42. R. Craig, G. Robinson, and P. Hatfield, "Performance of High Temperature SCR Catalyst System at Unocal Science and Technology Division," paper 92-109-08, in Ref. 15.

43. D. L. Champagne, *Gas Turb. World*, 20–23 (Nov.–Dec. 1987).

44. B. K. Speronello, J. M. Chen, and R. M. Heck "Family of Versatile Catalyst Technologies for NO_x and CO Removal in Co-Generation," paper 92-109.06, in Ref. 16.

45. D. W. Agar and W. Ruppel, "Extended Reactor Concept for Dynamic DeNO_x Design," *Chem. Eng. Sci.* **43**(8), (1988); J. R. Kittrell, C. W. Quinian, and J. W. Eldridge *J. Air Waste Manage. Assoc.* **41**(8), 1129–1133 (Aug. 1991).

46. T. D. Hylton, *Environ. Prog.* **11**(1), 54–57 (Feb. 1992).

47. Y. Wang, H. Shaw, and R. J. Farrauto, "Catalytic Oxidation of Trace Concentrations of Trichloroethylene over 1.5° Platinum on alpha-Alumina," in Ref. 3.

48. T. C. Yu, H. Shaw, and R. J. Farrauto, "Catalytic Oxidation of Trichloroethylene over PdO Catalyst on Alpha Al₂O₃," in Ref. 3.

49. J. L. Lin and B. E. Bent, "Thermal Decomposition of Halogenated Hydrocarbons on a Cu (111) Surface," in Ref. 3.

50. S. L. Hung and L. D. Pfefferle *Environ. Sci. Technol.* **23**(9), 1085–1091 (1989).

51. M. Mirghanbari, D. J. Muno, and J. A. Bacchetti, *Chem. Proc.* **45**, 14 (Dec. 1982).

52. T. G. Otchy and K. J. Herbert "First Large Scale Catalytic Oxidation System for PTA Plant CO and VOC Abatement," paper presented at the *85th Annual Meeting & Exhibition, AWWMA*, June 1992, p. 2026.

53. K. J. Herbert, "Catalytic Oxidation of the Vent Gas from a PTA Plant," 1992.

Ronald L. Berglund
The M. W. Kellogg Company

EXPECTORANTS, ANTITUSSIVES, AND RELATED AGENTS

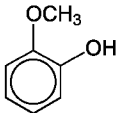
Expectorants

Expectorants enhance the production of respiratory tract fluid and thus facilitate the mobilization and discharge of bronchial secretions. Historically, expectorants have been divided into two classes based on specific mechanisms of action. Stimulant expectorants increase respiratory tract secretion by a direct effect on the bronchial secretory cells. Sedative expectorants act by gastric reflex stimulation. Many compounds classed as expectorants have been inadequately studied and the mechanisms of action are not known with certainty.

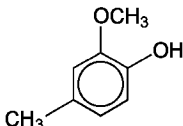
The clinical effectiveness of expectorants is a topic of significant controversy. The controversy results from a lack of accepted test methods for evaluating expectorants, with a consequent lack of significant objective data to support effectiveness. For most expectorant products, effectiveness is primarily based on a long history of use and a widespread subjective impression. However, the clinical effectiveness of one expectorant, guaifenesin, is supported by objective data. A randomized, double-blind, placebo-controlled study in 40 patients suffering from chronic bronchitis accompanied by productive cough was submitted to the FDA after this expectorant was classified by the FDA's Advisory Review Panel on Over-the-Counter (OTC) Cold, Cough, Allergy, Bronchodilator, and Antiasthmatic Drug Products as Category III, ie, available data insufficient to classify as safe and effective. Statistical data analysis convinced the FDA that guaifenesin loosens and thins sputum and bronchial secretions and makes expectoration easier by increasing sputum volume and reducing sputum viscosity (1). As a result, guaifenesin is the only expectorant permitted for use in U.S. over-the-counter products.

Expectorant preparations have changed dramatically since the 1930s with respect to composition and characterization. For example, in 1941 numerous expectorant formulas were described that contained 20 or more ingredients (2). One such formula, which was recommended for the relief of cough resulting from the common cold, contained thyme herb, horehound herb, grindellia, yerba santa, wild cherry bark, bloodroot, lobelia, squill, pleurisy root, life-everlasting, pipsisewa, mullein, comfrey, elecampane, ammonium chloride, menthol, eucalyptol, gaduol, cascarine, oil of white thyme, glycerol, honey, chloroform, alcohol, and sugar. In the 1990s only a few of the early natural expectorants are in widespread use, and some of these have been chemically modified to improve efficacy or physical characteristics.

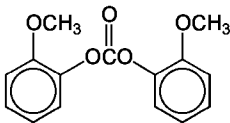
Guaiacols. Cresote, obtained from the pyrolysis of beechwood, and its active principles guaiacol [90-05-1] (1) and cresol [93-51-6] (2) have long been used in expectorant mixtures. The compounds are usually classed as direct-acting or stimulant expectorants, but their mechanisms of action have not been well studied. Cresol is obtained by the Clemmensen reduction of vanillin (3), whereas guaiacol can be prepared by a number of methods including the mercuric oxide oxidation of lignin (qv) (4), the zinc chloride reduction of acetovanillone (5), and the diazotization and hydrolysis of *o*-anisidine (6).



(1)



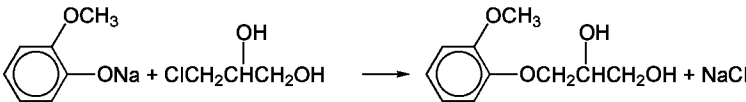
(2)



(3)

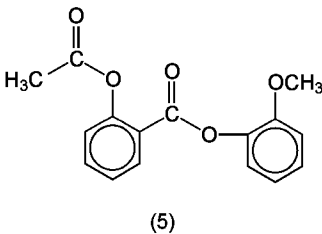
Because of its bitter taste and water insolubility, guaiacol has been chemically modified to improve its properties. Sulfonation provides a mixture of guaiacol-4- and 5-sulfonic acids which, as the potassium salts, is water-soluble, comparatively tasteless, but less active than guaiacol. Treatment of the sodium salt of guaiacol with phosgene provides guaiacol carbonate [553-17-1] (3) which also lacks the bitter taste of guaiacol, but is less water-soluble.

Guaifenesin [93-14-1] (4), formerly known as glyceryl guaiacolate, is the synthetic guaiacol derivative that has received the greatest acceptance as an expectorant. This compound is widely used, both in single-entity cough preparations and in combination with other active ingredients. Clinical studies carried out in the early 1940s indicated the usefulness of guaifenesin as an expectorant (7,8). Guaifenesin has also been shown to significantly increase the rate of clearance of inhaled radioactive particles in patients having chronic bronchitis (9). It is the only expectorant permitted for use in U.S. over-the-counter products by the Code of Federal Regulations final monograph on expectorant products (1). Guaifenesin may be prepared by the coupling of sodium guaiacolate and glyceryl monochlorohydrin.



(4)

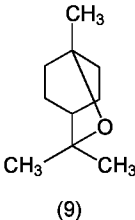
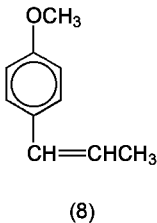
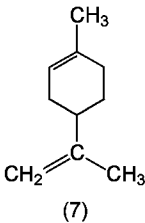
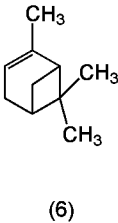
Guacetal [55482-89-8] (5), the acetylsalicylic acid ester of guaiacol, has been shown to retain both antiinflammatory and expectorant activity (10). It is used in Italy for symptomatic relief of painful respiratory disorders.



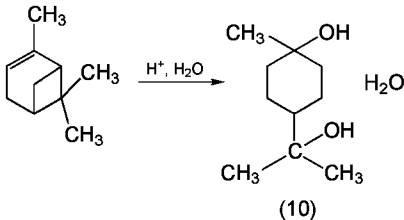
Volatile Oils. The use of volatile, sometimes called essential, oils as expectorants has been traced back 2000 years to Pliny who used turpentine internally to relieve coughing (11) (see OILS, ESSENTIAL). In spite of long usage, very few objective studies have been carried out to determine their real effectiveness. Limited evidence suggests that the compounds act by direct stimulation of the bronchial secretory cells. Compounds in this category were administered by a number of different routes including oral, topical, by aerosol, as a cream, and sometimes in lozenges. Volatile oils are not permitted for use in U.S. over-the-counter products by the Code of Federal Regulations final monograph on expectorant products (1). Menthol and camphor, however, are permitted in U.S. over-the-counter topical antitussive products, eg, lozenges, chest rubs, vaporizer additives (12).

The volatile oils are isolated from plant sources and are terpenoid in structure. They are purified by a combination of physical and chemical processes. Individual components of the oils are often isolated by crystallization or, in some cases, prepared synthetically.

Turpentine, one of the most familiar oils to be used in expectorant formulations, is prepared by the process of rectification, which consists of steam distillation of crude turpentine from sodium hydroxide, to remove the acidic and resinous components. The rectified oil contains primarily α -pinene [80-56-8] (6) along with small amounts of β -pinene. Two common household oils formerly used as expectorants are oil of lemon and oil of anise. Oil of lemon contains about 90% limonene [138-86-3] (7) and oil of anise, obtained by steam distilling the dried ripe fruit of *Pimpinella anisum* L., contains primarily anethole [104-46-1] (8). Oil of eucalyptus was first introduced into European medicine during the eighteenth century, but received the greatest attention during the mid-nineteenth century when it was used as a substitute for cinchona alkaloid antimalarials (see ALKALOIDS; CHEMOTHERAPEUTICS, ANTICANCER). During this period the expectorant properties were recognized. Oil of eucalyptus, prepared by steam distilling the leaves of *Eucalyptus globulus* Labill., generally contains not less than 70% cineole (eucalyptol) [470-82-6] (9).

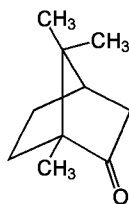


Terpin hydrate [2451-01-6] (10), one of the most well-known expectorants, is isolated from crude pine rosin left after the distillation of volatile terpene hydrocarbons and alcohols. It is also manufactured from turpentine (α -pinene) by acid-catalyzed hydration. Terpin hydrate may exist as cis and trans isomers, but only the cis isomer forms a stable, crystalline monohydrate. Terpin hydrate is available in the United States only in prescription products.

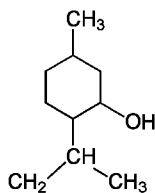


Camphor [126-04-5] (11), menthol [89-78-1] (12), and thymol [89-83-8] (13) are used in topical over-the-counter cough and cold preparations. Camphor is isolated from the camphor tree, *Cinnomomum camphora* T. Nees & Eberneier, or prepared synthetically from α -pinene or isoborneol. About 75% of the camphor sold in the United States is synthetic. Menthol, commercially the most important terpene alcohol, is obtained by crystallization from

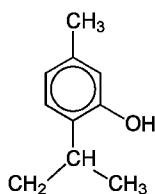
peppermint oil or prepared synthetically in racemic form by the hydrogenation of thymol. Menthol's local anesthetic activity may contribute to its antitussive properties. Thymol, unlike the other terpenoids described, contains a fully aromatic ring. It is obtained from the essential oils of *Thymus vulgaris* L. or can be prepared synthetically from *p*-cymene or *m*-cresol.



(11)

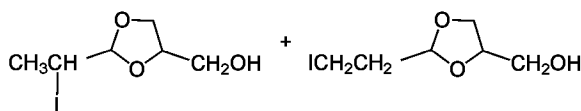


(12)



(13)

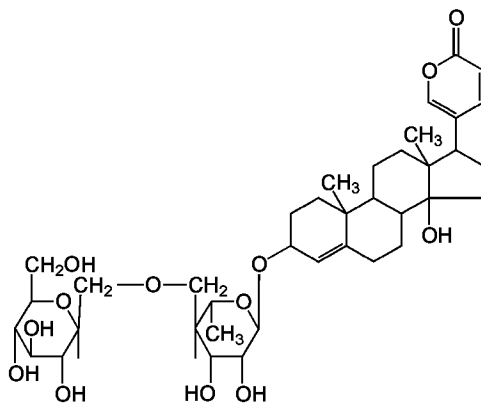
Iodides and Other Inorganic Compounds. Inorganic compounds such as potassium iodide [7681-11-0], hydriodic acid [10034-85-2], antimony potassium tartrate [28300-74-5], and ammonium chloride [12125-02-9] are thought to act by gastric reflex stimulation. Of these, only the iodides have been studied to any appreciable extent (13). A number of toxic reactions have been associated with both antimony potassium tartrate (14) and inorganic iodides (15,16). Reaction of iodine with glycerol [56-81-5] (qv) produces a stable organic iodide mixture, iodinated glycerol [5634-39-9] (14), which has expectorant properties. Formulations of (14) contain no free or ionic iodine, but iodine is released metabolically.



(14)

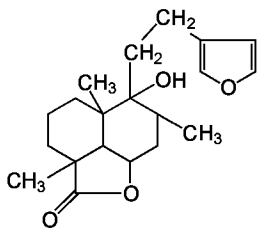
Miscellaneous Natural Products. The long list of natural products used in early cough preparations has diminished to the point where only a few are still mentioned. Four natural products still used occasionally outside the United States are squill, horehound, cocillana, and ipecac.

Squill is the dried, sliced bulb of *Urginea maritima* L. Baker and contains the glycosides scillaren A [11003-70-6] (15) and scillaren B [1393-22-2], the mixture of glycosides remaining after the separation of scillaren A. It has long been used in cough preparations and, oddly enough, red squill is also used as a rat poison. As an expectorant, squill is usually administered in combination with other ingredients. It is an emetic when given in large doses and this property suggests that the expectorant effect is likely to be caused by reflex stimulation. Toxic effects include nausea, vomiting, and a digitalis-like action on the heart (17).

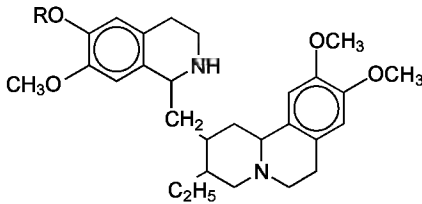


(15)

Horehound is the dried leaves and flowering tops of *Marrubium vulgare* L. Bickerman. It is reported that Serritus, in an early sixteenth century article on cough remedies, recommended the use of syrup of horehound to "purge through the sputa" (18). Although it has frequently been used in cough lozenges, little has been reported concerning its efficacy. The active principle in horehound is the diterpene lactone, marrubiin [465-92-9] (16).



(16)



(17) R = CH₃
(18) R = H

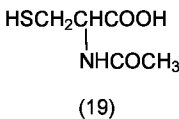
Ipecac is prepared from the dried roots and rhizomes of *Cephaelis ipecacuanha* (Brot.) A. Rich. and contains the alkaloids emetine [483-18-1] (17) and cephaeline [483-17-0] (18) in a ratio between 2:1 and 4:1. It has been used extensively in cough preparations and is believed to act by gastric reflex stimulation. Toxic effects include vomiting, irritation of the gastrointestinal tract, and cardiac arrhythmias (19). Ipecac syrup is available over-the-counter in the United States only in 30-mL containers for use as an emetic in treating poisonings.

Cocillana, the dried bark of *Guarea rusbyi* (Britt.) Rusby, was probably first used by the natives of the Bolivian Andes as an emetic–cathartic. It is often prescribed as an alternative to ipecac in the treatment of cough, and the emetic side effects at high doses suggest a mechanism of action similar to that of ipecac.

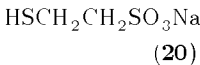
Mucolytics

Mucolytics reduce the viscosity of tenacious and purulent mucus, thus facilitating removal. The distinction between mucolytics and other classes of expectorants is frequently blurred. Steam, sometimes in conjunction with surfactants or volatile oils, has long been used to decrease viscosity by physical hydration. However, agents that chemically depolymerize certain components of mucus are available. Trypsin and other proteolytic enzymes have shown good clinical activity because of their ability to cleave glycoproteins. Pancreatic dornase, which depolymerizes DNA found in purulent mucus, also has shown clinical utility.

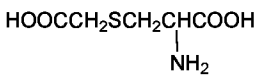
Several mucolytics reduce the viscosity of mucus by cleaving the disulfide bonds that maintain the gel structure. *N*-Acetyl-L-cysteine [616-91-1] (19), introduced in 1963, and mesna [19677-45-5] (20), developed in Europe in the early 1970s (20,21), are effective compounds in this class. Whereas most mucolytics must be administered by aerosol, carbocysteine [638-23-6] (21), which contains a derivatized sulfhydryl group, has shown activity by the oral route (22,23). However, carbocysteine does not reduce mucus viscosity, as does acetylcysteine, but appears to have a direct action on mucus glycoprotein production (24).



(19)

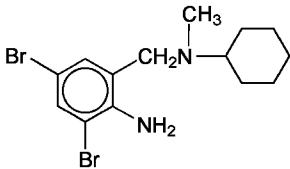


(20)

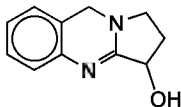


(21)

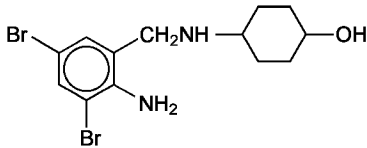
Bromhexine [611-75-6] (22), a highly substituted aniline derivative, has been shown to decrease the viscosity and increase the volume of mucus. Part of its activity has been attributed to its ability to fragment mucopolysaccharide fibers. Bromhexine is structurally related to the alkaloid vasicine [6159-55-3] (23), the active principle from the plant *Adhatoda vasica* Nees which has been used by the East Indians as an herbal medicine for the relief of cough. The pharmacological and clinical properties of bromhexine have been reviewed (25,26). The preparation of bromhexine follows standard procedures to *N*-(2-aminobenzyl)-*N*-cyclohexyl-*N*-methylamine which is then brominated to give (22) (27).



(22)

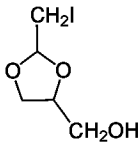


(23)

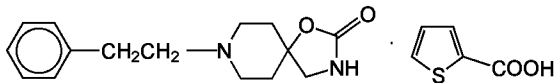


(24)

Ambroxol [18683-91-5] (24), a metabolite of bromhexine, has also shown potent clinical activity (28). Various esters of ambroxol have been shown to be 1.1 to 1.6 times more active as expectorants than ambroxol (29). The esters also show better gastric tolerability and more rapid absorption than ambroxol. *cis*-4-Hydroxymethyl-2-iodomethyl-1,3-dioxolane [61508-55-2] (25), a compound structurally similar to one component of iodinated glycerol (14), was shown to exhibit expectorant activity and mucolytic activity of the same order of magnitude as carbocysteine in rabbits (30). It also had a direct liquefying action on mucus *in vitro*. Acute and chronic toxicity in rats, mice, and dogs was low. The compound was prepared by the reaction of *cis*-4-hydroxymethyl-2-bromomethyl-1,3-dioxolane and sodium iodide. Decasilate [76652-72-7] (26) has been shown to possess significant mucolytic activity in rabbits. It also reduces bronchial seizures induced by citric acid in guinea pigs (31).



(25)



(26)

Antitussives

Records from antiquity describe the use of soothing syrups and herbal extracts to control excessive coughing. The use of goat's milk fresh from the udder for acute and chronic cough was reported in a review of ancient Hebrew medicine (32). Galen, in the second century AD, may have been the first to report a truly effective antitussive preparation, camphorated opium tincture (18). Through the centuries, cough remedies have remained a popular item and are found in most medicine cabinets. Over 300 prescription and over-the-counter preparations are available (33), and retail sales of nonprescription cough syrups, expectorants, and cough drops in 1995 are expected to exceed \$2 billion (34).

The Cough Reflex. Coughing is a protective reflex and is one of several important mechanisms for clearing the respiratory tract of excessive secretions and foreign debris, thus ensuring normal gas exchange and minimizing infection. It can be described as occurring in three phases: (1) a short deep inspiration of air into the lungs; (2) compression of the air by closure of the glottis, and contraction of the thoracic, abdominal, and diaphragmatic muscles; and (3) rapid expulsion of the air when the glottis opens. The intensity of the cough varies according to the force behind the expiration. In some cases the expired air moving through the trachea may achieve a linear velocity of >800 km h⁻¹ (35). The cough reflex can be triggered by chemical, mechanical, or other stimuli to the sensory nerve endings of the respiratory tract. The mechanical receptors generally are confined to the mucosa of the large airway of the upper respiratory tract, especially the lower third of the trachea. The chemical receptors are distributed widely and respond to almost any irritant. Stimulation of the sensory receptors causes impulses to pass along afferent nerve pathways to the cough center in the medulla. Here they are coordinated and transmitted by efferent or motor pathways to abdominal and intercostal muscles, and the diaphragm. Evidence suggests that irritation of the bronchial mucosa may not lead directly to stimulation of the cough receptors. Instead, bronchoconstriction may result first, which then triggers the cough reflex (13,36). The intensity of the cough is regulated to some extent by the stretch receptors located in the alveolar walls of the lungs.

Coughing can be produced by a number of environmental factors and pathological disorders. The common cold is the most frequent cause of transient cough in children and adults, and cigarette smoking is the most common cause of chronic, persistent cough (37). Changing causative factors from the 1870s to the 1970s are summarized (38).

Other factors that affect cough and the expulsion of irritants from the respiratory tract include ciliary activity and the production and physical characteristics of respiratory tract fluid. Under normal circumstances, the sweeping action of the cilia propels the demulcent fluid secreted by the goblet cells and the bronchial glands toward the glottis where it is swallowed or expectorated. Respiratory tract fluid is polymeric by nature, containing numerous polysaccharide units attached to a protein core. Large amounts of water and salt can be bound within the matrix, providing an ideal medium for transporting bacteria and debris away from the lungs. Pathological conditions or other factors which alter either ciliary activity, or the composition, amount, and physical properties of respiratory tract fluid may affect the frequency and productivity of coughing. Excellent reviews and monographs dealing with mucociliary clearance (39) and the role and physical characteristics of respiratory tract fluid (40–42) are available.

Theoretically, the act of coughing can be affected directly or indirectly by one or more of the following: (1) elimination of the ultimate cause of the cough by treating the responsible pathological condition, or by removing the anatomical or environmental irritant; (2) raising the threshold for stimulation of cough peripherally by anesthetizing sensory nerve endings in the respiratory tract; (3) interruption of the sensory impulses to the medulla; (4) specific depression of the cough center in the medulla; (5) interruption of conduction along the motor pathways; (6) nonspecific depression of the central nervous system; and (7) facilitating bronchial drainage and mucociliary clearance.

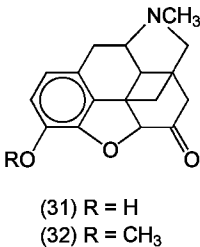
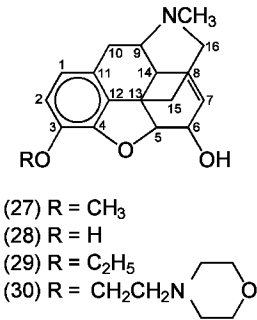
The first approach is considered ideal if the necessary relief can be obtained quickly enough. Most therapeutic agents, however, act by one or more of mechanisms (2), (4), and (7). The first pharmacological technique that permitted an objective evaluation of antitussive activity in animals was described in 1938 (43). Since that time, a number of methods have been developed that produce experimental cough by stimulating the respiratory tract, the vagus nerve, or the medulla by chemical, mechanical, or electrical means. Unfortunately, the large number of response criteria, species variation, and other factors reduce the value of these methods for predicting clinical efficacy and potency. Meaningful comparisons of antitussive drugs against pathologic cough in humans are often difficult to obtain owing to varied etiologies and intrinsic variability associated with small patient populations. For this reason, artificially induced cough is frequently used to measure antitussive activity clinically. A number of diverse chemical agents, such as citric acid, acetylcholine, acetic acid, sulfur dioxide, and ammonia have been used as stimulants. Several excellent reviews describe methods for the evaluation of antitussive drugs in animals and humans (44–46).

Centrally Active Antitussives. Centrally active antitussives depress the medullary cough center, thus raising the threshold for sensory cough impulses. The most well-known compounds in this category are the narcotics. Unfortunately, many of them have the disadvantage of addiction. Molecular modifications of the morphine skeleton and the synthesis of totally new structures have produced more specific drugs without the disadvantage of addiction. Many of the synthetic compounds also possess other useful properties including local anesthetic and antispasmodic activity (see also ANALGESICS, ANTIPYRETICS, AND ANTIINFLAMMATORY AGENTS).

Narcotic Antitussives. Since its isolation in 1832, codeine [76-57-3] (27) has been one of the most widely used and effective compounds for the treatment of cough. Though less potent than morphine [57-27-2] (28), it has become the reference against which most antitussives are measured.

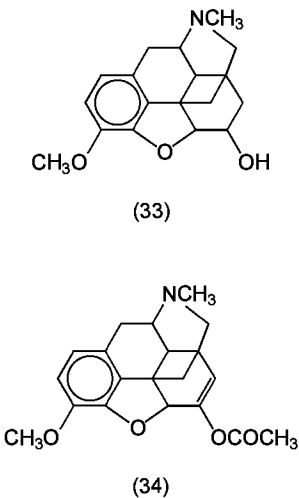
Codeine, like morphine, is isolated from the opium poppy. However, the low yield of 0.7–2.5% does not provide sufficient material to meet commercial demands. The majority of marketed codeine is prepared by methylating the phenolic hydroxyl group of morphine. Morphine yields from opium poppy are 4–21%. When prescribed for cough, the usual oral dose is 10–20 mg, three to four times daily. At these doses, adverse side effects are very few. Although the abuse potential for codeine is relatively low, the compound can substitute for morphine in addicts (47).

Molecular modifications of the morphine skeleton have produced numerous derivatives with antitussive properties, some of which have become commercially significant. Ethylmorphine [76-58-4] (29), a simple homologue of codeine, is prepared by ethylating morphine. It is pharmacologically similar to codeine but is seldom used clinically. Pholcodine [509-67-1] (30), the morpholinoethyl derivative of morphine, is used as an antitussive in a number of European countries. It is about one and a half times as potent as codeine, has little or no analgesic activity, and produces minimal physical dependence. The compound is prepared by the aminoalkylation of morphine (48).

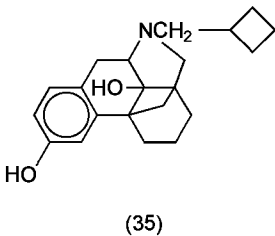


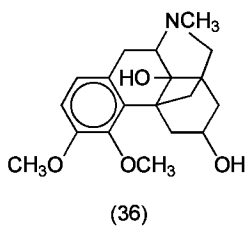
Hydromorphone [466-99-9] (31) and hydrocodone [125-29-1] (32) are isomers of morphine and codeine, respectively. Hydromorphone can be prepared by catalytic rearrangement of morphine (49) or by oxidation of the aliphatic hydroxyl group of dihydromorphine (50). Hydrocodone can be similarly prepared. As an antitussive, hydromorphone is several times more active than morphine and hydrocodone is slightly more active than codeine. Hydromorphone has a much higher addiction potential than hydrocodone.

Dihydrocodeine [125-28-0] (33), introduced in Germany before 1930, and dihydrocodeinone enol acetate [466-90-0] (34) both have clinical activity and addiction potential comparable to codeine.

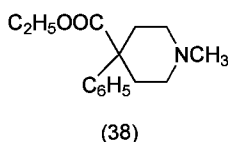
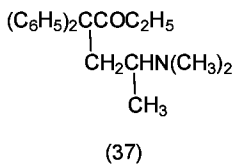


Modifications of the morphine skeleton have produced butorphanol [42408-82-2] (35) and drotebanol [3176-03-2] (36), which in animal models have demonstrated antitussive activity much greater than that of codeine (51,52). Butorphanol is also a potent analgetic of the narcotic antagonist type (51). Both compounds possess a unique 14-hydroxyl group.



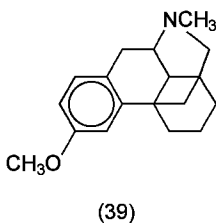


Among the nonopiate narcotics, two compounds, methadone [1095-90-5] (37) and meperidine [57-42-2] (38) have shown antitussive activity. Methadone is qualitatively morphine-like, and has demonstrated clinical activity against coughing induced by ammonia (53) and citric acid aerosol (54). Meperidine (38), although normally not thought of as an antitussive, has been shown to inhibit spasmodic coughing associated with bronchial asthma (55). The addiction potential of both of these agents limits clinical use for treating cough. A convenient synthesis for methadone which avoids the production of undesirable isomeric intermediates has been described (56). Meperidine can be synthesized by a number of different routes, but the method starting with phenylacetonitrile is one of the most practical (57).



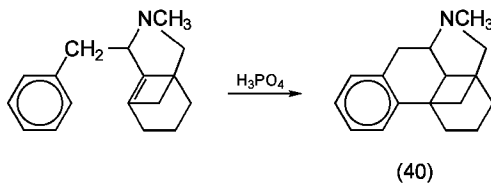
Side effects associated with narcotics include nausea, anorexia, and constipation; most of them also diminish ciliary activity and produce a drying effect on the respiratory tract mucosa.

Nonnarcotic Antitussives. The most centrally active, nonnarcotic antitussive is dextromethorphan [125-71-3] (39). It is similar to codeine in terms of potency and mechanism of action, ie, it is a direct depressant of the cough center. It is unique in that even though it is structurally related to codeine, it is not addictive.

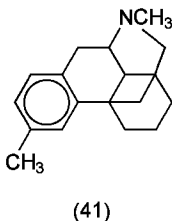


The synthesis of dextromethorphan is an outgrowth of early efforts to synthesize the morphine skeleton. *N*-Methylmorphinan(40) was synthesized in 1946 (58,59). The 3-hydroxyl and the 3-methoxy analogues were prepared by the same method. Whereas the natural alkaloids of opium are optically active, ie, only one optical isomer can be isolated, synthetic routes to the morphine skeleton provide racemic mixtures, ie, both optical isomers, which can be separated, tested, and compared pharmacologically. In the case of 3-methoxy-*N*-methylmorphinan, the levorotatory isomer levorphanol [77-07-6] (levorphan) was found to possess both analgesic and antitussive activity whereas the dextrorotatory isomer, dextromethorphan (39), possessed only antitussive activity. Dextromethorphan, unlike most narcotics, does not depress ciliary activity, secretion of respiratory tract fluid, or respiration.

The Grewe synthesis of *N*-methylmorphinan [3882-38-0] (40), which paved the way for the preparation of dextromethorphan and numerous analogues, follows standard reactions to 2-methyl-1-benzyl-1,2,3,4,5,6,7,8-octahydroisoquinoline. Cyclization of this compound with phosphoric acid gave a mixture of isomers from which *N*-methylmorphinan was separated.



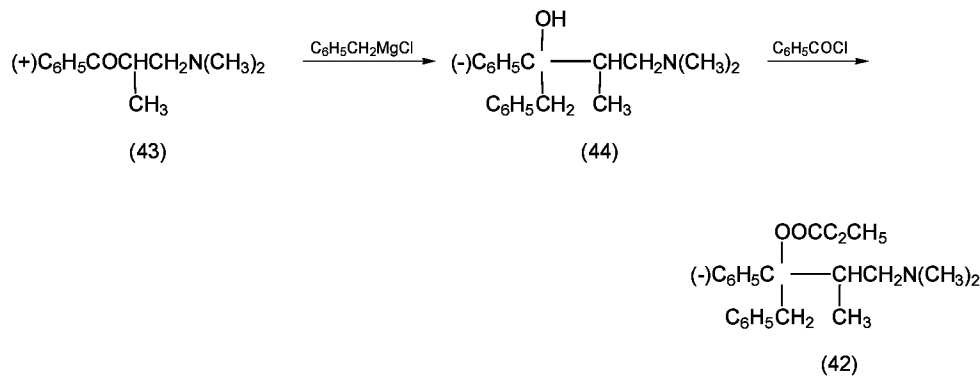
The synthesis (60) and potent antitussive activity (61) of dimemorfan [36309-01-0] (41), *D*-3-methyl-*N*-methylmorphinan, have been reported. This compound, prepared by a modification of the Grewe process, differs from dextromethorphan only by having a methyl group, rather than a methoxy group, in the 3 position.



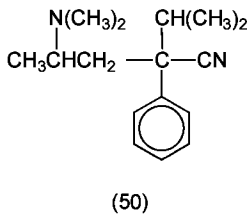
Included in the nonnarcotic class of antitussives are many compounds that do not possess a morphine skeleton and which vary widely from each other with respect to structural features and pharmacologic profiles.

Levopropoxyphene [2338-37-6] (42), the optical antipode of the dextrorotatory analgetic propoxyphene, is an antitussive without analgetic activity. The 2-naphthalenesulfonate salt has a less unpleasant taste than the hydrochloride salt, and is widely used. Clinical effectiveness has been demonstrated against pathological and artificially induced cough, but the potency is somewhat less than codeine. The compound is reported not to cause addiction. Levopropoxyphene can be prepared (62) by first resolving β -dimethylamino- α -methylpropiofenone with dibenzoyl-(+)-tartaric acid. The resolved

(+)-propiophenone [93-55-0] (43) is then treated with benzylmagnesium chloride to give (44), which is converted by acylation to levopropoxyphene.

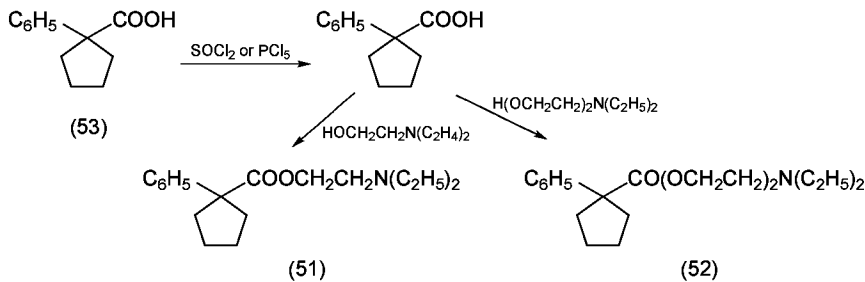


</

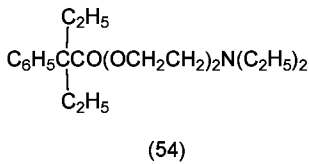


The citrate salt of isoaminile [77-51-0] (50) is a nitrile used as an antitussive in numerous European countries. In clinical trials it was shown to be as effective as codeine or chlrophedianol, with few mild side effects. Isoaminile citrate is longer acting than chlrophedianol and does not cause the respiratory depression of codeine (68).

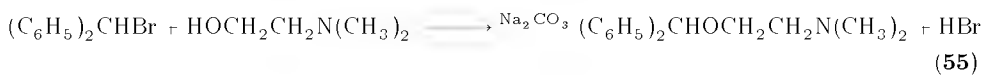
Caramiphen [125-86-0] (51) and carbetapentane [77-23-6] (52) are structurally related antitussives which, like chlrophedianol, also possess antispasmodic and local anesthetic activity. Caramiphen is also a potent anticonvulsant (69) (see HYPNOTICS, SEDATIVES, AND ANTICONVULSANTS). These compounds differ in that the ester side chain of carbetapentane is lengthened by an ethyleneoxy unit. Caramiphen is usually administered for antitussive activity as an ethanedisulfonate salt. It is comparable to codeine in potency but appears to have a longer duration of action. Experimentation has provided evidence that caramiphen has a central rather than a peripheral site of antitussive action (70). Carbetapentane is slightly less active than codeine. Both caramiphen and carbetapentane produce typical atropine-like side effects including dryness of the mouth and visual disturbances. They can be synthesized (71,72) from a common intermediate, 1-phenylcyclopentanecarboxylic acid [77-55-4] (53).



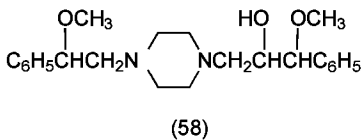
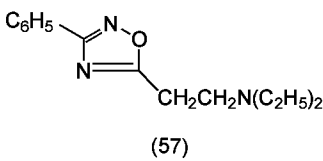
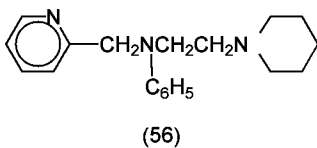
Oxeladin [468-61-1] (54), an antitussive developed in the UK, is structurally related to carbetapentane in that the cyclopentane ring has been broken at the 3,4-bond. It is also similar pharmacologically. Extensive clinical studies using cough drops and syrup containing oxeladin are described (73). The compound can be synthesized from phenylacetone nitrile (74).



Diphenhydramine [58-73-1] (55) was originally developed as an antihistamine and was first used clinically for this purpose in 1946 (see HISTAMINE AND HISTAMINE ANTAGONISTS). In addition to this primary effect, however, central antitussive activity has also been demonstrated in animals (75,76) and in humans (77). Its antitussive activity is about half that of codeine. Drowsiness is the most frequent side effect. Diphenhydramine can be prepared as follows (78):



Another antitussive with weak antihistaminic activity is the Japanese compound picoperine [21755-66-8] (56). This compound is a structural isomer of the well-known antihistamine tripeleminamine and is more potent than codeine. The chemistry (79) and pharmacology (80) of picoperine have been reported.

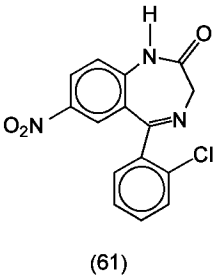
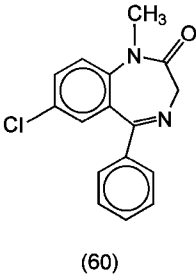
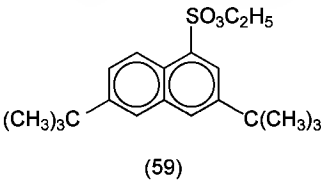


Oxolamine [959-14-8] (57) is sold in Europe. It is an oxadiazole, and its general pharmacological profile is described (81). The compound possesses analgesic, antiinflammatory, local anesthetic, and antispasmodic properties, in addition to its antitussive activity. Although a central mechanism may account for some of the activity, peripheral inhibition of the cough reflex may be the dominant effect. The compound has been shown to be clinically effective, although it is less active than codeine (82,83). The synthesis of oxolamine is described (84).

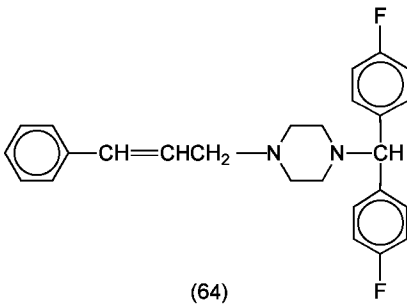
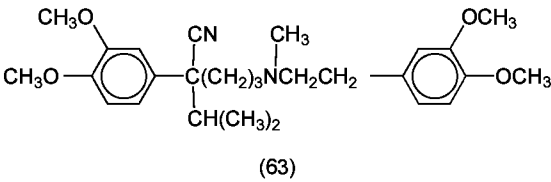
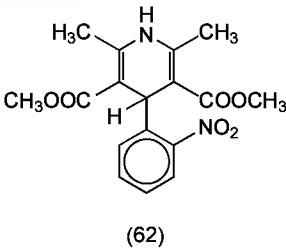
Zipeprol [34758-83-3] (58) is another European antitussive with a wide range of pharmacological effects, including antispasmodic, antihistaminic, and local anesthetic activities (85,86). It has been reported that zipeprol has been abused in Italy because high doses cause hallucinations (87). Spontaneous withdrawal symptoms similar to those of opiates have been observed; withdrawal symptoms can also be precipitated by naloxone. Zipeprol can be

prepared from 1-(2-methoxy-2-phenylethyl)piperazine and 3-methoxy-3-phenylpropylene oxide (88).

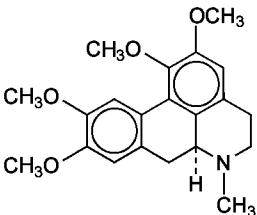
Ethyl dibunate [5560-69-0] (59), which is sold in Canada, is the ethyl ester of 3,6-(*tert*-butyl)-1-naphthalenesulfonic acid. It is structurally unrelated to most of the classical antitussives and is a selective central inhibitor of the cough reflex. Also significant is its low toxicity. The oral LD₅₀ is greater than 5000 mg kg in the rat. The clinical and pharmacological profile of this compound has been reviewed (89).



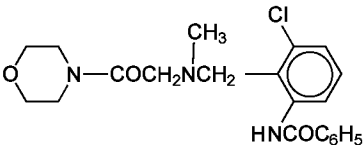
Diazepam [439-14-5] (60) and clonazepam [1622-61-3] (61) suppress cough induced by electrical stimulation of the lower brainstem of cats (90). Clonazepam and diazepam administered intravenously are about thirty-five times and six times more potent than codeine, respectively. Nevertheless, the compounds have not been widely used as antitussives in humans. Diazepam is used in the treatment of anxiety, and clonazepam as an anticonvulsant. The calcium ion antagonists nifedipine [21829-25-4] (62), verapamil [52-53-9] (63), and flunarizine [52468-60-7] (64) exhibit antitussive effects in a dose-dependent manner in guinea pigs (91). Pretreatment with a subthreshold dose of nifedipine also markedly increased the antitussive effects of morphine, dihydrocodeine, and dextromethorphan. However, none of the calcium ion antagonists are used clinically as antitussive agents. They are used in the treatment of angina and hypertension (see CARDIOVASCULAR AGENTS).



Glaucine [475-81-0] (65) has been compared against codeine for antitussive and toxic effects (92). In cats and dogs, the antitussive effect was 0.2 to 0.8 that of codeine, depending on the route of administration. The respiratory and cardiovascular effects of glaucine are similar to those of codeine but are generally less marked. Other studies concluded that glaucine appears to be safe for clinical use as an antitussive (93) and that the drug appears to exert its effect directly on the cough center (94).

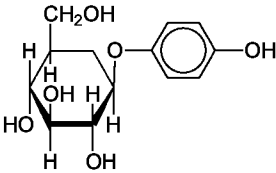


(65)

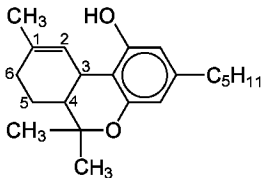


(66)

Fominoben [18053-31-1] (66) is another nonnarcotic drug which has shown antitussive activity comparable to codeine when administered both orally or parenterally in a variety of animal species (95).
Arbutin [497-76-7] (67) has been suggested as a potential antitussive (96). Both oral and intraperitoneal administration of arbutin to mice in a range of 50 to 200 mg kg showed a concentration-dependent activity against ammonia induced cough.



(67)



(68)

Among the long list of diverse structures reported to possess central antitussive activity is Δ^1 -tetrahydrocannabinol (THC) [1972-08-3] (68), the principal psychoactive component of marijuana (see PSYCHOPHARMACOLOGICAL AGENTS). This compound was found to be comparable to codeine against electrically induced cough in the anesthetized cat (90). Two other naturally occurring cannabinoids, cannabidiol and cannabinol, are inactive.
Except for the addiction liability of some of the narcotic antitussives, side effects for most of the centrally acting compounds are relatively few and mild at therapeutic doses. Qualitative comparisons of both side effects and pharmacological profiles have been summarized for many of the compounds described above (97).
Peripherally Active Antitussives. Peripherally active antitussives act by raising the threshold for cough at the sensory nerve endings, or by facilitating bronchial drainage, mucociliary clearance, or both. Agents that act by the first mechanism include local anesthetics that desensitize the mucosa of the respiratory tract, and antispasmodics that relax the smooth muscles of the bronchi (see NEUROREGULATORS). Most of these compounds also have a central antitussive component as a part of their pharmacological profile and have been previously discussed. Expectorants and mucolytics act primarily by the second mechanism.

Economic Aspects

Sales figures for expectorants and antitussives are usually combined under the general headings of cough preparations or cough and cold preparations. Antitussives and expectorants are frequently formulated together for the treatment of cough, or formulated with antihistamines and bronchodilators for the treatment of cold symptoms. Department of Commerce figures for the total value of manufacturer's shipments of cough or cough and cold preparations are shown in Table 1 (98–100).

Table 1. Shipment Values for Cough or Cough–Cold Preparations, \$ × 10⁶

Product	1967	1977	1991
cough preparations and expectorants ^a			
narcotic	27.4	56.7	133.1
nonnarcotic	15.9	40.2	30.1
cough–cold combinations ^a	7.3	11.6	132.0
cough–cold preparations ^b			
cough syrups	44.5	84.0	365.2
capsules and tablets	38.3	98.6	409.5
lozenges	14.8	25.8	51.1
topical preparations	0.9	⋮	⋮
cough drops		⋮	90.6
other preparations	23.1	98.8	224.3

^a Prescription products.
^b Over-the-counter products.

^c Shipment value combined with value for other preparations.

At the retail level, sales of nonprescription cough syrups, elixirs, and expectorants at all outlets were \$365.2 million in 1991 compared to \$304.8 million in 1978 (100,101). Sales of prescription cough preparations and expectorants were \$163.1 million in 1991 (100). Among narcotic antitussives, codeine (27) is by far the leading product. However, sales of codeine-containing cough preparations have declined in recent years, probably because of the increased acceptance of nonnarcotic antitussives such as dextromethorphan (39). U.S. figures show drugstore and hospital purchases of codeine for antitussive use declining from 7.71 t in 1972 to 6.38 t in 1977, with a projected decline to 5.77 t in 1981 (102).

Health and Safety

Safety and efficacy data on a number of antitussives and expectorants have been reviewed by the FDA's Advisory Review Panel on Over-the-Counter (OTC) Cough, Cold, Allergy, Bronchodilator, and Antiasthmatic Products. The conclusions and recommendations regarding the effectiveness, safety, labeling, and suitability for marketing of over-the-counter preparations have been reported (103). After review of these recommendations, FDA has issued final monographs for over-the-counter antitussives (12) and for expectorants (1). LD₅₀ data for most of the compounds described have been reported (104,105).

BIBLIOGRAPHY

"Expectorants, Antitussives, and Related Agents" in *ECT* 3rd ed., Vol. 9, pp. 542–560, by W. J. Welstead, Jr., A. H. Robins Co.

1. *Fed. Reg.* **54**(38), 8494 (1989).
2. E. J. Belanger, *Drug and Specialty Formulas*, Chemical Publishing Co., Inc., New York, 1941, p. 166.
3. J. H. Fletcher and D. S. Tarbell, *J. Am. Chem. Soc.* **65**, 1431 (1943).
4. **U.S. Pat. 2,433,227** (Dec. 23, 1947), H. F. Lewis and I. A. Pearl (to Sulfite Products Corp.).
5. **U.S. Pat. 3,057,927** (Oct. 9, 1962), D. W. Read (to Ontario Research Foundation).
6. **Ger. Pat. 1,148,236** (May 9, 1963), G. H. Herbst (to Hoechst AG).
7. W. J. Connell, G. M. Johnston, and E. M. Boyd, *Can. Med. Assoc. J.* **42**, 220 (1940).
8. W. F. Perry and E. M. Boyd, *J. Pharmacol. Exp. Ther.* **73**, 65 (1941).
9. M. L. Thompson, D. Pavia, and M. W. McNicol, *Thorax* **28**, 742 (1973).
10. C. Daroda, C. Galluzzo, R. Porzia, E. Riva, and M. Giovanni, *Minerva Med.* **72**, 363 (1981).
11. E. M. Boyd and G. L. Pearson, *Am. J. Med. Sci.* **211**, 602 (1946).
12. *Fed. Reg.* **52**(155), 30042 (1987).
13. H. Salem and D. M. Aviado, eds., *International Encyclopedia of Pharmacology and Therapeutics*, Vol. I, Pergamon Press, Ltd., Oxford, UK, 1970, p. 236.
14. S. C. Harvey, in L. S. Goodman and A. Gilman, eds., *The Pharmacologic Basis of Therapeutics*, Macmillan Publishing Co., Inc., New York, 1975, p. 928.
15. R. M. Harden, *Brit. Med. J.* **1**, 160 (1968).
16. G. E. Steffen, *J. Am. Med. Assoc.* **192**, 571 (1965).
17. A. Wade, ed., *Martindale The Extra Pharmacopoeia*, 27th ed., The Pharmaceutical Press, London, 1977, p. 604.
18. H. A. Bickerman, *Clin. Pharmacol. Ther.* **3**, 353 (1962).
19. Ref. 17, p. 601.
20. S. N. Steen and co-workers, *Clin. Pharmacol. Ther.* **16**, 58 (1974).
21. M. Tekeres, A. Horvath, L. Bardosi, and P. Kenyeres, *Clin. Ther.* **4**, 56 (1981).
22. M. Aylward, *Curr. Med. Res. Opin.* **2**, 387 (1974).
23. A. Quevauvillier and co-workers, *Therapie* **22**, 485 (1967).
24. C. R. Livingstone, M. A. Andrews, S. M. Jenkins, and C. Marriott, *J. Pharm. Pharmacol.* **42**, 73 (1990).
25. A. P. Launchbury, in G. P. Ellis and G. B. West, eds., *Progress in Medicinal Chemistry*, Vol. 7, Appleton Century Crofts, New York, 1970, p. 44.
26. Ref. 13, Vol. III, p. 765.
27. J. Keck, *Ann.* **662**, 171 (1963).
28. S. Pöschmann and R. Engelhorn, *Arzneim. Forsch.* **28**, 887 (1978).
29. **U.S. Pat. 4,528,393** (July 9, 1985), A. Moroni (to Magis Farmaceutici Srl).
30. G. Cantarelli, M. Carissimi, P. Gentili, and F. Ravenna, *Farmaco. Ed. Prat.* **34**, 393 (1979).
31. M. Ucelay, L. Labeaga, A. Orjales, L. Zubiaur, and A. Quintana, *Arzneim.-Forsch.* **41**, 528 (1991).
32. B. L. Gordon, *Ann. Med. Hist.* **4**, 219 (1942).
33. H. A. Bickerman, in W. Modell, ed., *Drugs of Choice 1976–1977*, C. V. Mosby Co., St. Louis, Mo., 1976, p. 446.
34. V. Cardinale, ed., *Drug Topics*, Medical Economics Co., Oradell, N.J., Feb. 4, 1991, p. 28.
35. E. M. Greisheimer, *J. Am. Med. Women's Assoc.* **15**, 1165 (1960).
36. H. Salem and D. M. Aviado, *Am. J. Med. Sci.* **247**, 585 (1964).
37. R. S. Irwin, M. J. Rosen, and S. S. Braman, *Arch. Intern. Med.* **137**, 1186 (1977).
38. Ref. 13, p. 271.
39. A. Wanner, *Am. Rev. Respir. Dis.* **116**, 73 (1977).
40. M. J. Dulfano, ed., *Sputum, Fundamentals and Clinical Pathology*, Charles C Thomas, Springfield, Ill., 1973.
41. E. Puchelle, ed., *Rheology of Bronchial Secretions and Respiratory Functions*, Masson and Co., Paris, 1973.
42. *Br. Med. J.* **2**, 51 (1975).
43. A. M. Ernst, *Arch. Int. Pharmacodyn. Ther.* **58**, 363 (1938).
44. Ref. 13, p. 205.
45. N. B. Eddy and co-workers, *Codeine and Its Alternatives For Pain and Cough Relief*, World Health Organization, Geneva, Switzerland, 1970, p. 127.
46. K. Bucher, *Agents Actions* **4**, 377 (1974).
47. C. K. Himmelsbach, *J. Am. Med. Assoc.* **103**, 1420 (1934).
48. **U.S. Pat. 2,619,485** (Nov. 25, 1952), P. Chabrier, P. R. L. Giudicelli, and C. H. Genot (to Laboratoires Dausse).
49. **Ger. Pat. 623,821** (Jan. 6, 1936), H. Metzger (to Knoll AG).
50. **U.S. Pat. 2,654,756** (Oct. 6, 1953), A. H. Homeyer and G. B. De la Mater (to Mallinckrodt Chemical Works).
51. R. L. Cavanagh, J. A. Gylys, and M. E. Bierwagen, *Arch. Int. Pharmacodyn. Ther.* **220**, 258 (1976).
52. S. Kobayashi and co-workers, *Arzneim. Forsch.* **20**, 43 (1970).
53. U. Trendelenburg, *Acta. Physiol. Scand.* **21**, 174 (1950).
54. H. A. Bickerman and S. E. Itkin, *Clin. Pharmacol. Ther.* **1**, 180 (1960).
55. A. L. Barach and H. A. Bickerman, *Pulmonary Emphysema*, Williams and Wilkins Co., Baltimore, Md., 1956, p. 168.
56. N. R. Easton, J. H. Gardner, and J. R. Stevens, *J. Am. Chem. Soc.* **69**, 2941 (1947).
57. F. F. Blicke and co-workers, *J. Am. Chem. Soc.* **74**, 1844 (1952).

58. R. Grewe, *Naturwissenschaften* **33**, 333 (1946).
59. R. Grewe and A. Mondon, *Chem. Ber.* **81**, 279 (1948).
60. M. Marakamu and co-workers, *Chem. Pharm. Bull.* **20**, 1706 (1972).
61. Y. Kase and co-workers, *Arzneim. Forsch.* **26**, 353 (1976).
62. A. Poland, L. R. Peters, and H. R. Sullivan, *J. Org. Chem.* **28**, 2483 (1963).
63. C. A. Winter and L. Flataker, *Proc. Soc. Exp. Biol. Med.* **81**, 436 (1952).
64. C. A. Winter and L. Flataker, *Toxicol. Appl. Pharmacol.* **3**, 96 (1961).
65. V. Giuliano and G. Rossa, *Minerva Med.* **46**, 1502 (1955).
66. **U.S. Pat. 2,778,824** (Jan. 22, 1957), C. Von Seemann (to American Home Products).
67. **Brit. Pat. 815,217** (June 17, 1959), R. Lorenz, R. Grosswald, and H. Henecka (to Bayer AG).
68. J. Diwan, R. Dhand, S. K. Jindal, S. K. Malik, and P. L. Sharma, *Int. J. Clin. Pharmacol. Ther. Toxicol.* **20**, 373 (1982).
69. F. C. Tortella, J. M. Witkin, and J. M. Musacchio, *Eur. J. Pharmacol.* **155**, 69 (1988).
70. E. F. Domino, H. Krutak-Krol, L. Jawahar, *J. Pharmacol. Exp. Ther.* **233**, 249 (1985).
71. **Swiss Pat. 234,452** (Jan. 16, 1945), (to J. R. Geigy AG).
72. **Brit. Pat. 753,779** (Aug. 1, 1956), H. G. Morren.
73. F. Kleibel, E. Steinhoff, and W. Wilde, *Therapiewoche* **15**, 1115 (1965).
74. **U.S. Pat. 2,885,404** (May 5, 1959), V. Petrow, O. Stephenson, and A. M. Wild (to British Drug House Ltd.).
75. J. Wax, C. V. Winder, and G. Peters, *Proc. Soc. Exp. Biol. Med.* **110**, 600 (1962).
76. N. B. Eddy and co-workers, *Bull. WHO* **40**, 682 (1969).
77. L. S. Lilienfield, J. C. Rose, and J. V. Princiotto, *Clin. Pharmacol. Ther.* **19**, 421 (1976).
78. **U.S. Pat. 2,421,714** (June 3, 1947), G. Rieveschl, Jr. (to Parke, Davis Co.).
79. Y. Kase and co-workers, *J. Med. Chem.* **13**, 704 (1970).
80. Y. Kase and co-workers, *Arzneim. Forsch.* **19**, 1916 (1969).
81. B. Silvesterni and C. Pozzatti, *Br. J. Pharmacol.* **16**, 209 (1961).
82. M. DeGregorio, *Panminerva Med.* **4**, 90 (1962).
83. K. Gary and W. T. Ulmer, *Arzneim. Forsch.* **17**, 240 (1967).
84. **Ger. Pat. 1,097,998** (Jan. 26, 1961), (to Firma Angelini Francesco).
85. G. Rispat and co-workers, *Arzneim. Forsch.* **26**, 523 (1976).
86. D. Cosnier and co-workers, *Arzneim. Forsch.* **26**, 848 (1976).
87. L. Janiri, P. Mannelli, A. M. Persico, S. Diodato, and E. Tempesta, *Drug Alcohol Depend.* **27**, 121 (1991).
88. **Ger. Pat. 2,109,366** (Sept. 30, 1971), Y. Roland and co-workers, (to Centre Europeen de Recherches Mauveray).
89. I. Shemano and J. M. Beiler, in Ref. 13, Vol. II, p. 615.
90. R. Gordon, R. J. Gordon, and R. D. Sofia, *Eur. J. Pharmacol.* **35**, 309 (1976).
91. J. Kamei and Y. Kasuya, *Eur. J. Pharmacol.* **212**, 61 (1992).
92. Y. Kase and co-workers, *Arzneim. Forsch.* **33**, 936 (1983).
93. M. I. Mironova and co-workers, *Farmakol. Toksikol.* **46**, 100 (1983).
94. Y. Kase and co-workers, *Arzneim. Forsch.* **33**, 947 (1983).
95. G. B. Fregnan and co-workers, *Drugs Exp. Clin. Res.* **8**, 601 (1982).
96. S. Li, G. Liu, Y. Zhang, and J. Xu, *Yaoxue Tongbao* **17**, 720 (1982).
97. Ref. 45, p. 239.
98. *Pharmaceutical Preparations, Except Biologicals, 1975*, Current Industrial Reports, Series MA-28G(75)-1, U.S. Bureau of the Census, Department of Commerce, Washington, D.C., 1976.
99. *1972 Census of Manufacturers, Industry Statistics*, Vol. II, Part 2, U.S. Bureau of the Census, Department of Commerce, Washington, D.C., 1973, p. 14.
100. *Pharmaceutical Preparations, Except Biologicals, 1991*, Current Industrial Reports, Series MA-28G(91)-1, U.S. Bureau of the Census, Department of Commerce, Washington, D.C., 1992.
101. B. P. Johnson, ed., *Drug Topics*, Medical Economics Co., Oradell, N.J., June 1, 1979, p. 27.
102. F. Streit and M. J. Nicolich, *Previous and Predicted Kilogram Purchases of Schedule II Drugs, 1972–1983*, Office of Medical and Professional Affairs, National Institute on Drug Abuse, Rockville, Md., 1979.
103. *Fed. Reg.* **41**(176), 38312 (1976).
104. M. Windholz, ed., *The Merck Index*, 10th ed., Merck & Co., Rahway, N.J., 1983.
105. E. J. Fairchild, ed., *Registry of Toxic Effects of Chemical Substances*, Vol. II, Dept. of HEW, Washington, D.C., 1977.

David B. Paul
Robert H. Dobberstein
Sandoz Pharmaceuticals Corp.

EXPERT SYSTEMS

At a basic level of description, expert systems are computer programs. However, they are different from most other computer programs in several ways. Functionally, they differ because they can perform problem-solving or decision-making tasks in some well-defined domain at a performance level comparable to human experts. In terms of content, expert systems primarily encode and manipulate symbolic knowledge about some domain or some type of problem-solving task, rather than mathematical equations, algorithms, or numerical data. Because of their emphasis on knowledge rather than on numerical computation, expert systems are often known, more appropriately, as knowledge-based systems. Knowledge-based systems are characterized by their dependence on fairly specialized knowledge, of the kind that humans accumulate through experience, understanding, or insight. By contrast to mathematical models, which describe physical and chemical phenomena, knowledge-based systems are qualitative models that capture human decision-making processes (1).

From this follows the definitions:

Definition 1: An expert system is a computer program that manipulates large amounts of symbolic knowledge using qualitative techniques, to solve problems that can otherwise be solved only by expert human problem solvers. Expert systems capture the human problem solver's expertise in the form of domain-specific knowledge and domain-independent problem-solving strategies.

Definition 2: Knowledge-based systems are computer programs that encode symbolic knowledge about domains and tasks, and solve problems by manipulating this knowledge using qualitative techniques.

The term domain means some subject or area of expertise. The domain can be as broad as chemical engineering, or it can be a more specific subject such as biochemical processes, distillation columns, or catalytic reactors. A problem-solving task is an intelligent reasoning activity that humans perform well. Examples include selecting among several alternatives and deriving conclusions based on known facts. Knowledge, as far as expert systems are concerned, is an aggregate of facts, descriptions, and methods expressed in the form of symbols and qualitative relationships.

Definition 2 is phrased in terms of knowledge-based systems rather than expert systems. No reference is made to expert human problem solvers. Definition 2 captures the sense that the representation and manipulation of knowledge is the source of such a system's power, whether or not that knowledge is directly elicited from a human expert.

Importance of Expert Systems. One way to understand what expert systems are is to understand what they do. Mycin, one of the earliest expert systems, is a computer program designed as a decision aid for doctors (2). When given data describing a medical patient's symptoms, Mycin could diagnose infectious blood diseases and prescribe therapies appropriate to the disease diagnosed. Another early knowledge-based system, R1, is still used today by Digital Equipment Corp. (3). Given a set of specifications describing the type of computer system required by a customer, R1 selects the appropriate computer components and peripherals, checks for inconsistencies, designs the layout of the entire system, and prints out a detailed order. The order can then be used by a sales representative to deliver what the customer specified.

These two examples hint at a few of the reasons for the importance of knowledge-based systems. A medical facility may handle hundreds of infectious disease cases a year. Speedy, accurate diagnosis of these cases, aided by a system such as Mycin, may help the medical facility handle more patients, more effectively. Likewise, configuring large computer systems composed of many components can be a time-consuming and error-prone task. R1 accomplishes this task correctly in a fraction of the time it would take for human technicians. The savings in this case are in terms of the number of orders processed, which ultimately translates to dollars. As a final example, consider Prospector, another classic expert system built in the 1970s (4). This computer program, designed to detect commercially viable ore deposits based on geological data, correctly identified a molybdenum ore deposit worth about \$100 million.

Other advantages of knowledge-based technology include the following: knowledge-based systems can automate complex tasks and do them quickly, even in fairly demanding domains, at a level comparable to their human counterparts; they can handle large amounts of data without suffering from information overload; they are tireless and potentially available round-the-clock; they are consistent, ie, if the inputs are the same, so are the outputs; they provide a readily available, lasting repository of expertise; and they have had proven successes in a variety of tasks, from diagnosis to design to data interpretation, and a variety of domains, from medicine to finance to chemical engineering.

Although knowledge-based systems have had many successes, there have been failures as well. Knowledge-based solutions may not be appropriate for certain problem-solving tasks. Alternative technologies may be more suitable, or the right approach to modeling the task may not yet have moved from the research lab to the shop floor. Failures have also happened for a variety of reasons associated with how the technology was applied and managed. Like some other technologies, expert systems have suffered from excessive hype and overblown expectations, both of which point to the need for care in appropriately applying the technology to the problem.

History. Knowledge-based system technology arose out of efforts to apply research in the field of artificial intelligence (AI) to practical problems. Broadly speaking, AI is concerned with trying to model intelligence and human problem-solving behavior using computational techniques. Central to this endeavor is the notion that computing machinery can emulate intelligence, a concept that was first formalized in 1950 (5). AI as a science is thus roughly 40 years old, and its early history makes fascinating reading (6). The evolution of AI can be divided into three segments, each characterized by certain key developments that led to knowledge-based systems:

The early period. In the 1950s and 1960s, AI was mostly concerned with developing computer programs that could perform tasks that were considered to require a high degree of intelligence, eg, game playing in domains such as chess and checkers (7); the solution of logic problems, such as theorem proving (8); and the solution of puzzle problems in toy domains (9). One key development of this period was the idea of heuristics, an important precursor to expert systems. Heuristics can be defined as guidelines for choosing among alternative courses of action. Heuristics can be used as shortcuts to direct the search for a solution along more promising lines, even if an optimal solution is not necessarily guaranteed. Another key development of this period was the creation of Lisp, a symbolic programming language.

The "understanding" period. The late 1960s and early 1970s saw AI begin to address broader aspects of intelligence. Research turned toward modeling cognition, interpreting natural language, story understanding, and ways to represent and reason about diverse kinds of knowledge (10–12). This period also saw the beginnings of applying AI research toward practical, real-world problems. Examples include Macsyma, which could solve mathematics problems symbolically (13); Strips, which could construct plans to achieve some goal, as an ordered series of steps (14); and Dendral, a program to assist in determining the molecular structure of chemical compounds (15). Programs such as these could perform at near expert levels in specialized domains, but still relied to some extent on search techniques rather than on a knowledge base.

The modern period. The significant breakthrough that led to the expert systems explosion of the late 1970s and early 1980s was the realization that computer programs could perform useful tasks at expert levels of performance, if they could be endowed with large amounts of specialized knowledge, and constrained to narrow but real domains. Instead of attempting to solve general problems of intelligence, research turned toward trying to clone human experts by capturing their experiential knowledge. Techniques were developed for extracting human expertise in the form of situation-specific rules, and encoding the rules into computer programs that could manipulate them and solve specific problems. Mycin, Prospector, R1, and other successful expert systems from this period were working proof of this concept. These successes led to the notion of an expert system shell, a skeleton program

that had the basic structure in which rules about a new domain could be entered, and the basic pattern matching capability to make inferences using the rules. The simplicity of the shell concept, combined with the potential for wide applicability, resulted in rapid commercialization. Starting with the early 1980s, people started applying knowledge-based systems to a wide variety of domains, including chemical engineering.

Knowledge-Based Systems in Chemical Engineering. In the chemical engineering domain, knowledge-based technology has been applied mainly to the tasks of diagnosis and design. One of the earliest applications was Falcon (Faults AnaLysis CONsultant), built jointly by Du Pont, The University of Delaware, and Foxboro (16). This was a rule-based system interfaced to on-line process data, that could automatically diagnose malfunctions in an adipic acid reactor unit. By the standards of the day, Falcon was a relatively large knowledge-based application. Its early success, combined with its on-line implementation, sparked interest in knowledge-based technology in chemical engineering.

Examples of diagnostic applications include Diad-Kit Boiler, an on-line expert system for performance monitoring and diagnosis of steam boilers (17), and CatCracker, a knowledge-based system for diagnosing operating problems in fluidized catalytic cracking units (18). Examples of design applications include BioSep Designer, which can automatically configure a process flow sheet and size the equipment involved in downstream separation of fermentation products (19); and Still, a tool for automated process and mechanical design of distillation columns (20).

Apart from design and diagnosis, other types of problem-solving tasks have also received attention in the chemical engineering domain, although not to the same degree. For example, knowledge-based reasoning techniques have been applied to simulation of discrete event processes (21), hazard and operability studies (22), intelligent multivariable process control (23), process operations planning (24), and scheduling of batch processes (25–27).

Technology Overview

From a technology perspective, knowledge-based systems (KBS) represent a new software methodology for solving certain types of problems effectively. It is important to understand what is encoded in knowledge-based systems, and how KBS technology differs from conventional numeric computational techniques.

Representation. From a software viewpoint, knowledge-based technology provides ways to represent knowledge and reason with it. The knowledge to be represented includes facts, descriptions, relationships, and problem-solving knowledge.

Facts About the Physical World. This category includes facts about specific objects or events, eg, reactor R-102 is refractory-lined, or the sun rose at 7:00 am yesterday. There may be generalized facts about the world, eg, grass is green, birds can fly, or copper sulfate crystals are blue. Assumptions on which these facts may rest, or conditions under which the facts no longer hold, are included in this category as well, eg, dry grass may be brown, or certain birds, such as penguins, cannot fly. Facts may also have associated with them the degree of certainty with which they are known.

Descriptions of Physical Objects, Processes, or Abstract Concepts. For example, pumps can be described as devices that move fluids. They have input and output ports, need a source of energy, and may have mechanical components such as impellers or pistons. Similarly, the process of flow can be described as a coherent movement of a liquid, gas, or collections of solid particles. Flow is characterized by direction and rate of movement (flow rate). An example of an abstract concept is chemical reaction, which can be described in terms of reactants and conditions. Descriptions such as these can be viewed as structured collections of atomic facts about some common entity. In cases where the descriptions are known to be partial or incomplete, the representation scheme has to be able to express the associated uncertainty.

Relationships Between Objects, Processes, and Events. Relationships can be causal, eg, if there is water in the reactor feed, then an explosion can take place. Relationships can also be structural, eg, a distillation tower is a vessel containing trays that have sieves in them; or relationships can be taxonomic, eg, a boiler is a type of heat exchanger. Knowledge in the form of relationships connects facts and descriptions that are already represented in some way in a system. Relational knowledge is also subject to uncertainty, especially in the case of causal relationships. The representation scheme has to be able to express this uncertainty in some way.

Problem-Solving Knowledge. The first three knowledge types are primarily declarative in nature, and most early approaches to knowledge-based systems generally focused only on these. In many cases, it is desirable to explicitly represent how-to knowledge, ie, knowledge about the reasoning strategies that are needed to manipulate the representations (28).

Reasoning. Knowledge-based systems need ways to manipulate their representations to produce useful results. Representation and reasoning are closely intertwined; how knowledge is represented influences and constrains the types of reasoning methods that can be brought to bear on the representations. Some fundamental techniques for reasoning are deduction, pattern matching, and search.

Deduction. If a knowledge-based system has a set of facts, and new facts are provided to it, then rules of inference can be applied to the set of facts to derive conclusions. For example, from the facts that (1) hydrogen and oxygen can react explosively at high temperatures, (2) air contains oxygen, (3) the atmosphere inside a particular reactor contains hydrogen, and (4) the reactor is at a high temperature, the additional fact that air has leaked into the feed to the reactor leads to the conclusion that an explosion can take place. The conclusion is based on applying deductive logic to the known facts. This is representative of the logic-based approach to knowledge-based systems.

Pattern Matching. If knowledge is expressed, not in terms of individual facts, but in terms of direct associations between situations and conclusions, then conclusions can be reached by trying to match the given situation with a stored conclusion. For example, consider a knowledge base with a number of associations relating types of service with the appropriate pump for that service. Assume that each association can be considered separately, ie, no interactions. Given a particular type of service, the task is to select the right pump. This selection can be made by trying to match the specifications of the current case with the preconditions of one or more associations in the knowledge base.

Search. In spite of the emphasis in KBS technology on knowledge being the source of problem-solving power, search is still one of the staple techniques used in manipulating representations. Search is closely associated with pattern matching capability. For example, if the pump selection knowledge base is very large, it would not be intelligent to try to match every association to the given specifications. Some means of constraining the pattern matching complexity would be required. A focused search, based on successively narrowing down the scope of the selection, would be one way to tame this complexity. Search techniques are also relevant in other tasks where the knowledge representation encodes solution alternatives or partial solutions.

Comparison to Conventional Numerical Programs. Fundamentally, knowledge-based systems manipulate symbols that have some meaning in the external world. For example, mathematical expressions, such as the one shown, can be integrated analytically, using the rules of integral calculus to manipulate the symbols representing the variables. This is how humans often solve integration problems.

$$\int (x^2 + 3x + 2) \, dx$$

By contrast, a numerical computer program for solving such integration problems would depend on approximating the mathematical expression by a series of algebraic equations over explicit integration limits.

Knowledge-based systems typically use qualitative methods rather than quantitative ones. For example, consider a simple tank system. The equation describing the flow rate of liquid out of the tank is given below, where C_i is the orifice coefficient, d is the diameter of the orifice, and h is the height of liquid in the tank. Based solely on the form of the equation, a human reasoner can infer that the flow rate F increases monotonically with the height h of liquid in the tank.

$$F = C_v \cdot d \cdot h^{1/2}$$

To reflect this type of reasoning, a KBS captures qualitative relationships between variables. By contrast, a conventional program that implements the flow equation calculates the value of the flow rate for numerical values of the input variables, ie, orifice diameter, orifice coefficient, and liquid height.

Humans deal robustly and efficiently with complexity by using shortcuts, guidelines based on past successes, and by relaxing the constraint of optimality. For example, in troubleshooting a process problem the human expert may use direct associations between causes and symptoms gathered from past experience. Similarly, knowledge-based systems make use of heuristic strategies for arriving at feasible solutions. There may even be more than one correct solution for the problem. By contrast, conventional numeric programs are usually oriented toward finding exact or optimal solutions. For example,

an optimization program that calculates operating targets might try to find exact values for the decision variables, so that an objective function is maximized subject to algebraic constraints.

Human cognitive tasks also have to cope with uncertainty. For example, real world constraints are subject to change without notice, one's knowledge may apply only partially to a particular situation, and the data needed to infer conclusions may be incompletely known. Knowledge-based systems deal with such problems by explicitly representing and reasoning with uncertainty. By contrast, conventional numeric techniques are applied to well-defined problems that can be mathematically characterized. Uncertainty is implicitly dealt with by making simplifying assumptions, or using empirically determined model parameters. The key difference is that the computational process does not deal with the uncertainty; it is the developer of the computational algorithm who does. These remarks are applicable primarily to deterministic numeric models. Stochastic models do explicitly address uncertainty by incorporating random distributions.

The types of data structures and processing techniques used in knowledge-based systems are different from those used in conventional, numerical programs. For example, knowledge-based systems use data structures such as rules and objects and process these using pattern matching, inheritance, and message passing. Often, knowledge-based systems are data-driven, ie, processing is initiated in response to changes in data, not in a predetermined sequence. By contrast, conventional computer programs use data structures that are more suited for representing, storing, and manipulating numbers, eg, arrays, records, fields, and tables. The types of processing techniques used include arithmetic operations, logical comparisons, iterations, and relational calculus. In general, processing tends to be program-driven, ie, a sequence of instructions acts upon data in a predetermined fashion.

Knowledge-based systems differ from conventional programs in the types of applications for which they are used. Knowledge-based technology is most appropriate for implementing tasks that have some cognitive equivalent. Thus KBS applications automate tasks such as troubleshooting, conceptual design, or scheduling, whereas conventional programs address such number-oriented tasks as data storage and retrieval, equation-solving, optimization, and numerical simulation. This is not to imply that numerical programs are of no use in performing tasks such as design or scheduling; only that they are not well-suited to the cognitive component of these tasks.

Theory

From the viewpoint of AI and cognitive science, the theoretical question underlying knowledge-based technology is how to model expert problem-solving behavior. The predominant models of representation and reasoning in AI are logic, rules, and objects. More recent approaches to knowledge-based systems include task-specific models. The emphasis here is on applying the theory, ie, how do the problem-solving models and the representation concepts apply to real-world problems? The representation and reasoning techniques are described independently of any specific development tool, and illustrated using examples from chemical engineering wherever possible.

Logic. The rule-based representation used in many expert systems has its roots in formal logic. Logic programming has been applied successfully to build knowledge-based applications for certain problems (29,30). It is a popular approach in Europe, and has been used in the much-touted Japanese Fifth Generation Computer Project. The logic approach supports a declarative style of representation; knowledge about a domain is represented in the form of facts or axioms. The fact base can then be queried about the truth or falsehood of some proposition to be tested. The logic-based system sets up the test proposition as a goal to prove, and uses rules of inference to do so. The most general of the inference rules is a method called resolution theorem proving (13).

Of the many variants of logic, one of the most important is first-order predicate logic. Facts in predicate logic are expressed using variables, constants, predicates, and quantifiers. For example, consider the facts shown below:

For all X such that Liquid (X) , Flow (X)

Liquid(Water)

In statements 1 and 2, "For all" is a quantifier, specifically, the universal quantifier; Water is a constant; X is a variable; and Liquid and Flow are predicates. Both statements 1 and 2 are asserted to be true. First-order predicate logic gets its name because predicates such as Liquid and Flow are allowed, and quantification is only allowed on variables, not predicates. Having represented these facts in the logic system, the system can then be queried:

Flow(Water) ?

The system would respond:

True

The logic system arrives at this answer by trying to prove the assertion in statement 3. The constant argument of the predicate Liquid is matched with any variable argument for the same predicate in other sentences in the fact base. The proof procedure is not as straightforward when the fact base contains a large number of sentences with different types of constructs. Sentences may have negations, eg, NOT(Corrosive(X)); conjunctions, eg, Corrosive(X) AND Viscous(X); disjunctions, eg, Corrosive(X) OR Viscous(X); existential quantifiers, eg, "For some X such that Acid(X), Corrosive(X)."

In such cases, simple pattern matching and variable binding will not suffice; the entire resolution theorem proving method has to be applied. For details, the reader is referred to more comprehensive texts (13,31).

First-order predicate logic is appealing from a mathematical standpoint because there are provable properties of logic-based systems (31). As a general methodology for implementing knowledge-based systems, however, first-order predicate logic has a number of practical shortcomings. There are restrictions on the expressiveness of facts because the syntax of assertions has to conform to the rigid framework of logic. There are no easy ways to structure or organize knowledge. Entities, events, processes, and the relationships between them have to be represented in a flat, one-dimensional space of facts. There are no straightforward ways to encode the uncertainty involved in assertions such as, "The presence of high concentrations of catalyst in the effluent may be indicative of poor settler operation." Systems based on first-order predicate logic have difficulties dealing with time. Neither is there any way of retracting conclusions that no longer hold true, or specifying defaults. Last, but not least, logic-based systems are notationally complex to use.

Extensions of first-order predicate logic have been developed that address some of these shortcomings. For example, fuzzy logic, first introduced in 1965 (32), addresses the problem of representing and reasoning with uncertainty. A number of temporal logics have been devised for representing time (33). Default logics and nonmonotonic logics have also been devised for addressing the problems of representing defaults and retracted conclusions (34). Problems of usability have been addressed well in programming environments for building logic-based systems. However, the addition of extensions to first-order predicate logic takes away many of the provable properties of such systems. This partly undermines some of the motivations of elegance and mathematical provability that prompted the use of logic systems in the first place.

At the implementation level, the Prolog language is probably the most popular software for developing logic-based expert systems. In the chemical engineering domain, the logic-based approach has not been applied extensively. In the United States, rule-based and object-based systems have enjoyed much greater popularity. In Europe, despite the popularity of Prolog and logic-based approaches in the knowledge-based systems community, there are few accounts of implemented systems in the chemical engineering literature (22).

Rules. Rules, first pioneered by early applications such as Mycin and R1, are probably the most common form of representation used in knowledge-based systems. The basic idea of rule-based representation is simple. Pieces of knowledge are represented as IF-THEN rules. IF-THEN rules are essentially association pairs, specifying that IF certain preconditions are met, THEN certain fact(s) can be concluded. The preconditions are referred to as the left-hand side (LHS) of the rule, while the conclusions are referred to as the right-hand side (RHS). In simple rule-based systems, both the

preconditions and conclusions are variable–value pairs. For example, a rule for diagnosing a reactor would be represented as follows.

IF reactor_temperature is high
THEN coolant_malfunction is established

Rule (1)

In this example, reactor_temperature and coolant_malfunction are variables, and high and established are values. Variables can take symbolic or continuous values. For example, the rule above can alternatively be stated as:

IF reactor_temperature is 150
THEN coolant_malfunction is established

Rule (2)

In general, both the preconditions and conclusions can have conjunctions (AND), disjunctions (OR), and negations (NOT). Conjunctions make a rule more specific, whereas disjunctions make a rule more general. A rule with a disjunction is equivalent to two separate rules with the same conclusion.

Figure 1 shows the architecture of a rule-based system. Fundamentally, there are three components: a rule base, a working memory, and an inference engine. The rule base contains discrete fragments of knowledge, all expressed as IF–THEN associations over variables. The working memory contains facts or assertions that are true at any stage in the computational process, ie, it contains variable–value binding. The inference engine is a domain and knowledge-independent processing mechanism. The inference engine searches the rule base, matching variable–value bindings in the working memory to preconditions and conclusions of rules in the rule base. When a match occurs, the rule is said to have fired. During this process, if a rule is traversed left to right, trying to infer conclusions from known conditions, the strategy is called forward chaining. If the rule is traversed right to left, trying to evaluate conclusions by checking if the appropriate conditions are met, the strategy is called backward chaining.

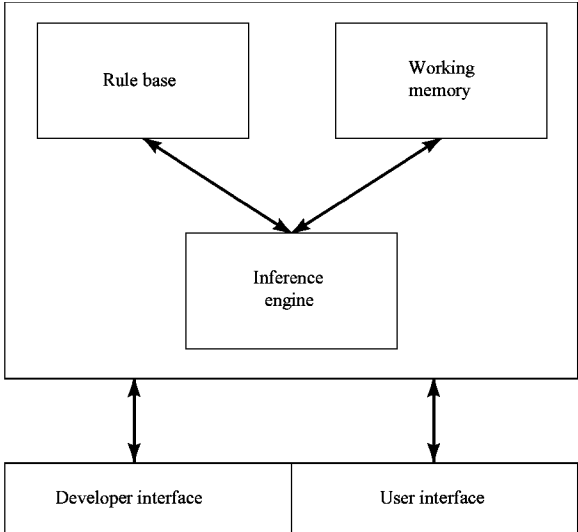


Fig. 1. Architecture of a rule-based system.

Forward chaining is useful when the objective is to find the implications of new pieces of information. This is applicable to tasks such as process monitoring, eg, determining the consequences of a change in process operating conditions, or design configuration, eg, determining how to sequence a series of separation steps. In forward chaining, the inference engine starts with some initial data, ie, values of variables in the working memory. It then fires every rule that applies to the starting data. The conclusions of the fired rules are also variable–value bindings that are introduced into working memory. The inference engine continues to apply rules based on the derived information until no more rules apply. An example of a forward chaining rule, used in a typical design configuration task, follows.

IF design_pressure is high
THEN configure safety_release_valve

Rule (3)

Backward chaining, on the other hand, is useful when the objective is to validate hypotheses. This is applicable to diagnostic and troubleshooting tasks, eg, finding the root cause of a process problem, or equipment selection tasks, eg, selecting a pump for a particular type of fluid service. In backward chaining, processing starts with the introduction of a goal into the working memory, ie, a conclusion to test. The inference engine tries to find all rules with the goal in their conclusions. Each rule that fires results in the precondition clause being added to the working memory as another goal. If data in the working memory can be bound to a goal, then that helps to validate or negate the goal. This process continues until no more rules fire. The reactor coolant malfunction rule (see Rule 1) is an example of a backward chaining rule used in a diagnostic task. The precondition of the rule is a symptom of a malfunction, while the conclusion represents a potential root cause. The objective of the system is to evaluate plausible conclusions by checking their preconditions against known data.

Forward and backward chaining can be combined in the same rule-based system to address subtasks in the overall problem. For example, the monitoring component of an on-line troubleshooting system can be designed to work in a forward chaining mode, while the component that determines root causes can be designed using backward chaining. When more than one rule is applicable in a given situation, the inference engine uses conflict resolution to decide which to fire. Several conflict resolution strategies can be used: (1) apply rules in the order in which they appear in the rule base (this is the most simplistic); (2) apply the most specific rule first; (3) apply rules according to a priority assigned at development time (static priority); and (4) apply rules according to a priority computed at run-time, based on other variables (dynamic priority).

Rules may represent either guidelines based on experience, or compact descriptions of events, processes, and behaviors with the details and assumptions omitted. In either case, there is a degree of uncertainty associated with the application of the rule to a given situation. Rule-based systems allow for explicit ways of representing and dealing with uncertainty. This includes the representation of the uncertainty of individual rules, as well as the computation of the uncertainty of a final conclusion based on the uncertainty of individual rules, and uncertainty in the data. There are numerous approaches to uncertainty within the rule-based paradigm (2,35,36). One of these approaches is based on what are called certainty factors. In this approach, a certainty factor (CF) can be associated with variable–value pairs, and with individual rules. The certainty of conclusions is then computed based on the CF of the preconditions and the CF for the rule. For example, consider the following example.

reactor_temperature is high with CF 80
IF reactor_temperature is high

(5)

THEN coolant_malfunction is established

Rule CF 90

(6)

Then the certainty of the conclusion, namely "coolant_malfunction is established" is computed as follows:

CF(conclusion) = CF(precondition) * CF(Rule) 100

= 80*90 100

= 72

(7)

Thus certainty factors express the intuitive notion that the certainty of a conclusion should be lower than the certainty of the data and knowledge involved in arriving at the conclusion. Certainty factor theory also allows for combining the CFs of conjunctions, disjunctions, and negations:

CF(NOT(A)) = 100 – CF(A)

(8)

CF(A AND B) = minimum[CF(A), CF(B)]

(9)

CF(A OR B) = maximum[CF(A), CF(B)]

(10)

The certainty factor approach has been among the more popular rule-based approaches to uncertainty. However, although it is easy to apply given the individual CFs, acquiring the raw CFs from the experts is often quite difficult. Further, although the formulas for CF combination are mathematically appealing, they often have no relation to the ways in which experts combine evidence to arrive at conclusions. Some of the task-specific approaches discussed later address uncertainty combination in a more intuitive way (35).

Rule-based systems address the issue of explaining conclusions as well. During the inference engine's processing, two types of explanatory questions can be answered: why and how. Why questions arise when the system makes requests for data in order to validate hypotheses. The system responds to this type of question by referencing the rule which has the data under question, eg, "The reactor_temperature is required because Rule 102 requires it to validate coolant_malfunction." Similarly, a rule-based system responds to the how question by tracing through all the rules that led to the final conclusion, and referencing them. In addition, rules can be associated with explanatory text, which the inference engine can retrieve to provide clearer explanations, eg, "The reactor_temperature is required because a high reactor_temperature may be indicative of a possible coolant_malfunction, which is the hypothesis currently under consideration in Rule 102. The knowledge for this rule was obtained from Jane Q. Expert."

Rules are clearly a useful form of representation for knowledge-based applications, with their advantages of representational simplicity, wide applicability, and history of past successes. However, certain important design criteria govern the proper application of rules and there are shortcomings of the rule-based representation.

Design Criteria. In any but the most simple systems, rule organization is important for problem-solving and system maintenance. Although individual rules may be robust, they may interact in strange and unexpected ways with other rules, if they are put into a large unstructured rule base. Grouping rules according to a common set of goals or contexts is important for this reason. For instance, an on-line troubleshooting system may have an overall rule organization based on the following subtasks: monitor the process, detect abnormalities, find the root cause, and recommend corrective actions. As far as possible, conflict resolution should be minimized, again by rule organization. The basic capabilities of the inference engine, ie, forward or backward chaining, may not be sufficient to capture the problem-solving strategy. Some higher level control of the problem-solving process, representative of the task being done, may be required. Such additional mechanisms have to be encoded as strategy rules, called meta-rules in the literature (37). Meta-rules control the application of the knowledge rules, or set the context in which the knowledge rules should be applied.

Shortcomings. Rule representations based solely on variable–value bindings are restrictive in their scope. For example, consider a knowledge base that contains rules for selecting rotary pumps for various fluid services. If it later turns out that some of the rules are also applicable to selecting positive displacement pumps, the original rule set must be modified to include positive displacement pumps using OR operators in the conclusions of those rules. Thus there is no easy way to generalize over classes of pumps. Secondly, there are other types of relationships essential to capturing knowledge about a domain that are outside the scope of a rule-based representation. Examples are "type-of" or "part-of" relations. The use of meta-rules, unavoidable in many cases, has the disadvantage of obscuring the actual problem-solving strategy used by the system. This is a problem for the developer. It is also a problem for the end user who is confronted with explanations that are far removed from the actual reasoning process. Providing clear explanations of problem-solving is a research area in itself (38,39).

Object-Oriented Representation. Object-oriented representation addresses many of the shortcomings of rules, especially when used in combination with a rule-based representation. In contrast to the type of data structures used in traditional programming, eg, variables, records, or linked lists, the basic data structure in object-oriented systems is an object with associated properties. For example, a process temperature measurement for reactor R-101 would be represented by the object R101_TEMPERATURE, with properties RAW_VALUE, SYMBOLIC_VALUE, and SYMBOLIC_TREND. Properties of an object can take values, eg, the SYMBOLIC_VALUE property of R101_TEMPERATURE might take the value HIGH. In many object-oriented systems, the object–property value holder is referred to as a slot, eg, R101_TEMPERATURE.SYMBOLIC_VALUE is a slot with a value of HIGH. Objects that share properties can be generalized into a class. For example, if there are multiple reactors, all reactor temperature objects could be grouped into a class called REACTOR_TEMPERATURES. In this sense, an object is also referred to as an instance of a class. Similar classes in turn can be grouped into higher level classes, ie, the REACTOR_TEMPERATURES class and other temperature measurement classes could be grouped together under TEMPERATURES. In turn, this and other classes, such as PRESSURES, can be grouped under the class PROCESS_DATA (Fig. 2).

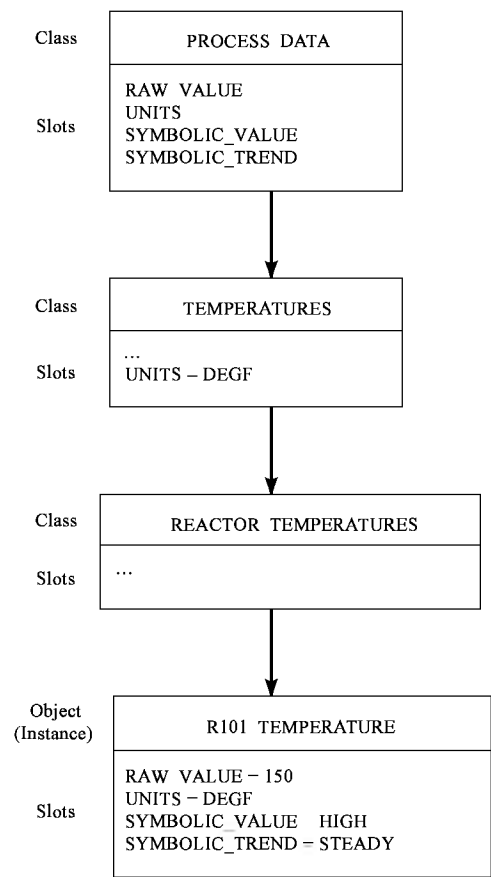


Fig. 2. Object-oriented representation.

Some object-oriented systems also support the notion of subobjects. This facilitates the representation of structural relationships. For example, the objects IMPELLER and LINING can be made subobjects of a REACTOR object, to represent the structural components of the reactor.

To illustrate the power of the object-oriented style of representation, consider the reactor diagnosis example used earlier in the discussion of rules. Assume that there are several reactors, R-101, R-102, etc, each served by a common cooling system. Relating coolant malfunctions to the temperature in each reactor would need multiple rules, or rules with multiple disjunctions. Instead, if rules are used in combination with the object representation described above, a single general rule can be written to cover all the specific instances, as follows.

IF REACTOR_TEMPERATURES.SYMBOLIC_VALUE is HIGH
 THEN coolant_malfunction is established

Rule (4)

In object-oriented systems, the notion of inheritance is important. Inheritance means that a property of a class is automatically acquired by subclasses and objects of that class. For instance, having defined all the classes and objects listed above, if the property `TIMESTAMP` has to be added to all process data objects, one simply has to add it to the class `PROCESS_DATA`. This modification would be automatically reflected in all subclasses and objects.

Another important idea in object-oriented systems is being able to specify how slots can obtain values and what types of values they can hold. This allows specification of data types for slots, initial and default values, and restrictions on value ranges. Also, procedures to execute for obtaining the value can be specified, or even procedures to execute after getting the value. These types of procedures are referred to as daemons. For example, in the objects belonging to the `REACTOR_TEMPERATURE` class, the `SYMBOLIC_VALUE` slot may change, so it would be convenient to call a procedure to calculate the symbolic value on an as-needed basis. A daemon can be placed in the slot to do this, based on current values in the `RAW_VALUE` slot. The `RAW_VALUE` slot can also be restricted to a certain range of values if it is known that reactor temperatures cannot physically exceed some upper bound. Daemons, and other information describing the values a slot may take, are entered into what is termed the facet of the slot.

These discussions address the issue of representation using objects, but what about reasoning? One approach is to couple the object representation with a rule-based system and let the associated inference engine take care of the reasoning. This is the technique used in a number of knowledge-based applications, and it successfully addresses many limitations of pure rule-based systems, eg, the reactor diagnosis example. However, some of the other restrictions of rule-based systems, namely limited problem-solving capabilities and the inadequacy of meta-rules, still remain. To address these, object-oriented systems offer powerful capabilities for implementing many different reasoning strategies, via encapsulation and message passing. Encapsulation allows procedures and data to be bundled together into objects. Rather than data being passive, and acted upon by procedures, encapsulation allows data to be active, and procedures to work differently depending on the object they are attached to. Daemons are an example of this. The second feature, message passing, allows objects to communicate. The overall reasoning process is accomplished as follows: objects send messages to other objects; the recipient objects respond by executing their associated procedures; results or values are returned to the sending objects. Using these advanced features of object systems requires the use of an integrated programming language such as Lisp or C. These types of capabilities are somewhat on the fringes of mainstream knowledge-based systems technology. They are more applicable in the area of model-based systems.

One of the many advantages of an object-oriented representation is the fact that it allows representation of all the declarative information known about a process or domain in a structured way. Representing the domain knowledge declaratively offers the flexibility to choose whatever reasoning strategy is appropriate for the problem to be solved. In addition, the idea of inheritance greatly facilitates building and maintaining systems using an object-oriented representation. On the downside, object systems do not generally come with built-in problem-solving strategies, uncertainty handling capability, or explanation features. This is up to the system developer to implement using a programming language. For a range of knowledge-based applications, this potential hurdle can be avoided by coupling the object system with a rule base. Many commercially available tools support this hybrid architecture. For other types of problems, the hybrid approach may not be sufficient, and the full power of object-oriented programming may have to be used. For example, knowledge-based simulation of discrete event processes can be approached in an object-oriented way (21).

Advanced Topics.

Model-Based Reasoning. Much of the knowledge in mainstream knowledge-based systems is associational in nature. Causal relationships, for example, are encoded as direct associations between symptoms and root causes, suppressing the details of any intermediate causal links. Knowledge in this form is also referred to as compiled knowledge (40). The advantage of this approach is that at problem-solving time, the knowledge is available to the system in the form needed. No explicit reasoning is required to derive the association. The downside is that if the knowledge-based system runs into a situation where those details become important, then it may fail. This is the brittleness problem. One way to address the brittleness problem within the

framework of mainstream knowledge-based systems, is to structure knowledge at successive levels of detail, so that additional detail is available when needed (18). Nevertheless, the details are still represented in compiled form, ie, they are not derived at problem-solving time. Ensuring that all the detail is represented explicitly as associations places some burden on the developer, and there are limitations on how much detail can be explicitly represented.

Model-based reasoning attempts to address the brittleness problem in a more fundamental way, by modeling the behavior of the important entities in the domain. The objective of this modeling effort is to enable the reasoning system to derive any associational knowledge at problem-solving time in a flexible way. For example, if the task is troubleshooting a chemical process, the fundamental entities are the process equipment and the unit operations that take place within them. The goal would be to qualitatively model both the normal and malfunctioning behaviors of the process equipment using first principles. Modeling reduces the burden of having to explicitly represent the behaviors of the entities in a compiled form, but requires additional development time. The advantage is that the modeling effort can be leveraged for future applications, eg, if generic models are developed for process equipment, then they can be re-used for other processes.

In general, the model-based reasoning approach is best applied, not as a method in itself, but as an add-on to a knowledge-based system. The main reason is that modeling is hard, and problem-solving based solely on fundamental models is computationally complex (41). Using the hybrid approach can take advantage of the efficiency of compiled knowledge in rapidly focusing on the solution, while retaining the robustness of models when confronted with the need for behavioral detail. Several recent research papers in the chemical engineering literature have explored this hybrid problem-solving notion (42,43). For more information on model-based reasoning in general see References 44 and 45.

Task-Specific Problem Solving. Rules and objects are examples of general problem-solving approaches, ie, the representations and reasoning constructs can be applied to different tasks. The work of mapping the needs of the task to the constructs available is left to the developer of the knowledge-based system. This can be a difficult knowledge engineering problem. For example, in the design configuration task, the constructs of importance are the components (or devices) to be included in the final configuration, the functions that those components can satisfy, and the design specifications to be met by the configuration. To build a knowledge-based configuration system, the developer has to map these task-level concepts to the basic constructs provided by rules, objects, or logic. Issues of how to appropriately represent the task-level knowledge, and how to structure the design rules in the rule base, for example, have to be resolved. Thus rules and objects can be regarded as fine-grained approaches to representation and reasoning, in the sense that they deal with problem-solving at a low level.

Task-specific problem-solving approaches address these issues of knowledge engineering, representational ease, and representational adequacy (28,46). Such approaches are oriented toward characterizing the representation and reasoning needs of different cognitive tasks in a generic way. The advantage is that if a knowledge-based system has to be constructed for such a task, the developer has a ready-made blueprint to work with. The blueprint specifies what representational constructs are required, what reasoning strategies are appropriate, what knowledge is required to build the application, and how to go about acquiring the knowledge. Some researchers in this area have also focused on building tools specific to the tasks, so that even the representation and reasoning methods are available to the developer in a prepackaged form (28).

Task-specific approaches have a number of advantages, eg, knowledge structuring to overcome brittleness, ease of development, knowledge acquisition, reusability, and ease of explanation. From a developer's point of view, the downside is that the problems to be solved have to be appropriately matched to known tasks, and if not, a new task characterization has to be done. From a research point of view, there are issues of granularity and knowledge sharing to be addressed: at what level of detail should the task be characterized? How can the same fundamental knowledge be applied to different tasks? Nevertheless, task-specific approaches have been successful in guiding knowledge-based system development. Several large-scale problems in the chemical engineering domain have been solved using this approach (18,20).

Real-Time Systems. Reasoning with time and having to solve problems under time constraints place unique burdens on a problem-solving system. To address these, knowledge-based systems need additional capabilities: problem-solving strategies to deal with the effect of time, increased processing speed, and enhanced integration capabilities, to name a few. Although real-time often means different things to different people, there is general agreement among researchers on the types of issues that real-time systems need to address (47,48). These are as follows.

- speed:* the system must be capable of completing its computations faster than the rate of arrival of fresh information
- responsiveness:* the system should be able to recognize and respond to events in the time scale of interest
- pre-emption:* the system must be capable of pre-empting ongoing tasks when interrupted by unscheduled events of greater importance
- adaptability:* the system must be able to change its behavior over time
- temporality:* the system must be able to represent time and comprehend temporal relationships among events

Of these, the first two place demands on the processing speed of the knowledge-based system and its integration with the environment, eg, data acquisition interfaces. The third impacts on the nature of the computational cycle and the system's ability to reschedule its own computations. The latter two place demands on the representation and reasoning capabilities of the system, and are the most interesting from a knowledge-based system viewpoint.

Temporal reasoning is nonmonotonic in nature, ie, conclusions that are made in a certain time step may have to be retracted in future time steps. Conclusions may have persistence, ie, once made, they may last for a certain period of time before being subject to change. Two concepts are of particular importance in temporal reasoning: state and event. A state is associated with some aspect or condition of the physical world, and may change with time within the time scale of interest. For example, high reactor temperature is a state of the process. It may change over time. On the other hand, an event is associated with a unique point in time. For example, a reactor runaway took place at 7:00 am on 7 2 89 is an event, unchanging with time. In most systems that do not operate in real time, states are of primary importance to the task, eg, the reactor diagnosis example used earlier in this article. In real-time knowledge-based systems, sequences of events provide additional evidence for making conclusions. All of these additional representational capabilities are needed for developing real-time knowledge-based systems (49). The representation and reasoning needs can either be addressed by making use of the rich, extensible features of object-oriented programming, or by extending the capabilities of rule-based systems. Some development tools provide built-in constructs to handle some of these representational needs, eg, G2 from Gensym Corp. There are few detailed accounts of implemented real-time knowledge-based systems in the chemical engineering domain. The Falcon system mentioned earlier is one example (16), but there are others (50–52).

Building Applications

The purpose of this section is to provide some understanding of where and how to appropriately apply knowledge-based technology, and to give an industrial perspective on the process of developing and delivering knowledge-based applications. The literature contains numerous detailed discussions concerning knowledge-based system development (53–55).

Technical and Business Issues. In general, the successful application of any technology is dependent on both technical and business issues. For knowledge-based systems, six critical issues are problem identification, user acceptance, measurement of impact, appropriateness, feasibility, and cost.

There should be an identifiable problem that requires a knowledge-based solution. Typically, problems amenable to a knowledge-based solution are identified by the existence of a knowledge bottleneck, ie, whenever the successful execution of a task is delayed, limited, or compromised by the speed or availability of an expert. For example, when process upsets occur during night shifts, the rapid resolution of the problem is hampered by the unavailability of an experienced process engineer. Alternatively, the rapid evaluation of a large number of design choices is often limited by the processing speed of the design team. A third type of identifiable problem is the routine task, which if automated would free up the human experts to expend their energies in more creative activity. An example of this is equipment selection, the automation of which would permit the design engineer to spend more time creatively exploring the design of the process flow sheet.

The importance of considering the potential user of the application cannot be emphasized enough. This is particularly true of knowledge-based systems because they usually involve some degree of human interaction. Most knowledge-based applications to date have been designed to serve in an advisory capacity, keeping the human decision maker in the loop. For example, even if the technical or operating departments at a manufacturing facility identify a potential knowledge-based system application, the ultimate user of the application has to endorse and accept the application and participate in its development.

The impact of a knowledge-based application may appear in many ways: improved competitive position, quality improvement, improvement in efficiency, cost reduction, and reduction in downtime, to name a few. Some of these benefits may be hard to quantify; others may not be quantifiable at all. For example, the actual benefit derived from a diagnostic advisory system may not be apparent if the process behaves normally. To quantify the benefits, a careful post-audit may have to be done, taking into account the number of faults averted, and comparing the frequency of faults before and after implementation.

The appropriateness (or otherwise) of knowledge-based technology may become apparent during problem identification. In general, knowledge-based systems are appropriate when the task involves some symbolic reasoning, the task uses at least some symbolic information, the domain of the task is narrow, there are human experts who can perform the task successfully in a few hours or a few days, and infallible performance is not a requirement. Knowledge-based systems, unless designed for graceful failure or augmented by other robust AI techniques, can potentially fail in a brittle fashion if completely novel situations are encountered or if the inputs are noisy and incomplete.

Alternative technologies should also be considered, and the reasons for not using them should be justifiable. For example, database technology is not the right choice if a task requires reasoning that goes beyond retrieval of stored data based on well-defined criteria. At the same time, many problems that are stated in a symbolic way can be formulated mathematically, and in fact can be better solved numerically. For such problems, knowledge-based systems are not the appropriate answer.

Even if knowledge-based systems are the correct choice, there may be potential roadblocks to successful implementation. For example, experts should be available and willing to be debriefed. There should be adequate support from management. End users should participate in the effort. There are technical issues such as the selection of the right tools for implementation. If knowledge-based technology is just one component of the overall solution (as it often is), other technologies which must be integrated should receive careful consideration. The required expertise for implementing knowledge-based solutions must also be available.

Most of the cost of building knowledge-based systems is associated with the time and effort involved in knowledge engineering, which is often an iterative exercise. Probably the best way to estimate this cost is by using past experience with similar projects as a basis. Common development times range from one to six work months for small systems, six months to a year for medium-sized systems, and one to two years for large systems. One approach to cut down this development time is by building knowledge bases in a reusable way. Apart from the cost of knowledge-based development, there may be development costs associated with integrating the knowledge-based system with other systems if needed, the cost of delivering and installing the application, and training and documentation. Other costs such as the license and annual maintenance cost of software used should also be factored in.

Approaches to Development. In general, there are two approaches to knowledge-based systems development: the do-it-yourself approach, and the technology approach. In the do-it-yourself approach, the experts, who are the knowledge sources, learn enough about the technology and the tools to construct their own knowledge-based systems. This approach is successful when applied to small, well-defined problems, but education in the technology is a critical prerequisite. The do-it-yourself approach is also well-suited to highly decentralized organizations, where the burden of development and deployment is often on local groups of users.

The technology approach is one in which knowledge-based system professionals interact with experts and end users to develop applications. This is often the only option if the application involves even moderate complexity, and low end expert system shells prove to be inadequate to represent the knowledge or the reasoning. Irrespective of the approach taken, the development process can be roughly broken down into the eight steps shown in Figure 3. The type of approach taken simply changes the relative importance of the steps. The only exception is that in the do-it-yourself approach, education may be a prelude to the eight steps discussed here. The experts or end users may have to undergo some training in the theory and application of knowledge-based technology, and the mechanics of using specific expert system tools.

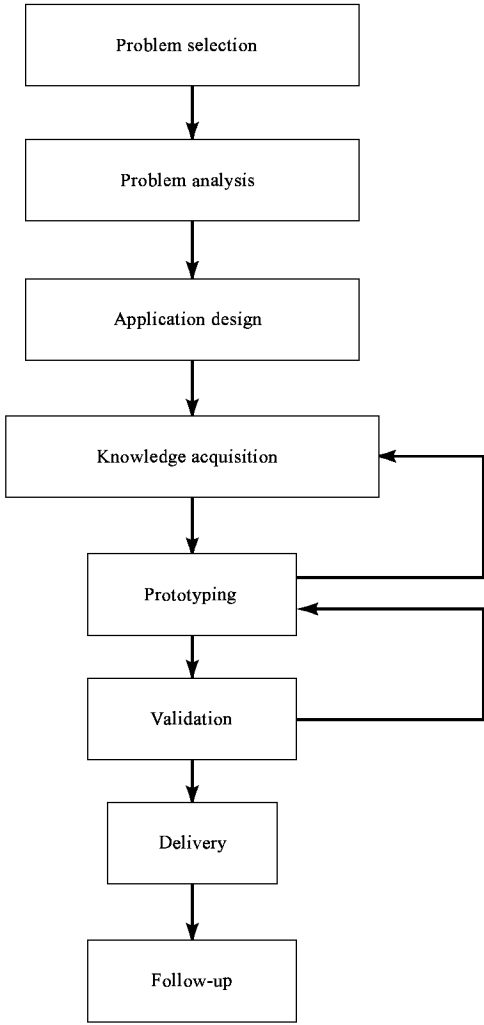


Fig. 3. Steps in the development process.

Problem Selection. To select the problem correctly, the criteria discussed earlier should be carefully applied before launching a project: the existence of a knowledge bottleneck, the inapplicability of exact numerical methods, the existence of either an expert or a theory for the task, the narrowness of the domain, and the business issues of payout and cost. If needed, the various criteria can be quantified and weighted based on their

importance to the problem in order to rate potential projects (56). If the application of knowledge-based technology is either purely exploratory or educational, this step may be dispensed with. The problem selection step will likely involve the end users, the experts, and management.

Problem Analysis. The characteristics of the problem affect the knowledge representation and reasoning requirements, which in turn impact on the tool used for development. Analyzing the characteristics of the task and the types of knowledge available can help identify these requirements. In many cases, this can be accomplished by protocol analysis, ie, a detailed step-by-step examination of actual problem-solving cases performed interactively with the experts(s). For example, the protocol analysis may reveal that the task consists of a well-defined sequence of decisions, with context-specific knowledge being brought to bear at each decision step. This may suggest a decision tree representation, with rules to evaluate each decision node. If the knowledge is available either as a collection of discrete facts or in the form of guidelines for actions, a rule-based or logic-based representation may be the appropriate choice. If the knowledge used in the task consists of relations such as "type of" or "part of," or if the rules can be generalized by applying them to classes of objects, then an object-oriented representation may be useful.

It is also useful to try and classify the problem into broad categories, such as diagnosis, design, selection, planning, or scheduling. This may help to map the problem directly into a predefined representation reasoning schema (57). If the task requires finding the best of many alternatives, then it is likely to involve search strategies. If the execution of the task is likely to involve many novel situations not generalizable a priori, then model-based reasoning or first principles reasoning may be called for. In many cases, there may be several components in the overall task. For example, diagnosis may involve data abstraction (interpretation of raw data), determining root causes, and selection of remedial actions (58). Problem analysis should be kept independent of domain details as far as possible.

Application Design. Application design involves fleshing out the functionality of the overall system. Rarely, if ever, is knowledge-based technology used in isolation in applications. The developer has to consider the integration of the knowledge-based component of the application with other components such as the user interface, data acquisition interface, and access to numerical routines for calculated data, as required. Figure 4 shows a typical application architecture that incorporates these components. End users and information systems staff are likely to be heavily involved in the application design phase, at least in a consulting capacity.

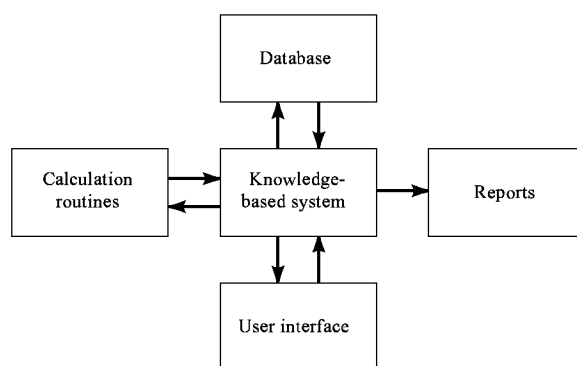


Fig. 4. Architecture of a typical integrated application.

Knowledge Acquisition. The process of knowledge acquisition is dependent on the results of problem analysis, ie, the representation and reasoning methods chosen for the problem guide the knowledge engineer through the elicitation (59). For example, knowledge acquisition for a diagnostic application would focus on the various root causes relevant to the domain, associations between root causes and symptoms, and ways to organize the space of root causes to better focus the search. However, there are a few general methods that are useful, independent of the task. These include protocol analysis, repertory grids, and induction from examples (36). Protocol analysis is a technique for obtaining case studies and following the detailed steps in the expert's chain of reasoning. Repertory grids are a way of organizing associations in the form of tables, noting differences between associations, or other attributes of the associations, eg, level of confidence. From an object-oriented point of view, repertory grids are particularly useful for identifying features of interest, and for assigning values to those features. Lastly, induction from examples is a useful approach when the representation is simple and relatively unstructured. Given actual problem-solving examples of data and the conclusions resulting from the data, an induction system can generalize a decision tree or a set of unstructured rules. Knowledge engineers and experts are the primary participants in the knowledge acquisition step.

Prototyping. As Figure 3 shows, the prototyping step is a parallel activity with knowledge acquisition, conducted in an iterative fashion. Unlike traditional software applications, knowledge-based systems are difficult to specify exactly during the problem analysis phase. Intermediate results from the prototyping step may help with further knowledge acquisition. The prototypes can help get feedback from the end users. Prototyping also results in a phased development effort; as modules of the system get developed, they can be tested and even used, while development on other modules progresses. This step requires a good understanding of representation techniques, reasoning methods, and the features of the development tool. Experts and knowledge engineers are the primary participants in this phase; however, end users may also be involved to provide feedback on system performance.

Validation. Validation by experts and end users drives the iterative process shown in Figure 3 for prototyping and knowledge acquisition. Validation is also concerned with how to test the system after the knowledge engineers and experts have completed their work and produced a fully functional application. This is an important but difficult activity. The performance of the system has to be measured, its robustness evaluated, and its areas of brittleness identified. Performance measurement, as used here, refers to measuring the accuracy of the knowledge-based system's conclusions, and the confidence level to be expected from the system. Other common performance measures include execution speed and response time. An obvious performance measurement is to test the application on available case studies, for which the inputs and results are known. However, this should not be the only method of validation used because available cases usually constitute a small fraction of the scope of the system. There are other validation methods that focus on the content and consistency of the knowledge in the system, checking for rule conflicts, redundancies, circular rules, and knowledge gaps (60).

Delivery. This includes installation and commissioning of the application, user training, manuals, and system documentation. Delivering knowledge-based systems is not very different from delivering other types of applications, with the exception that end users may need more time and assistance in getting familiar with using the system. It is important to keep the needs of delivery in mind during the early phases of the project, since they can have a profound influence on overall system design (not just the knowledge-based component), and can substantially impact time and cost estimates for projects.

Follow-Up. Even after the system is delivered, additional work is usually needed. In conventional software engineering this is viewed as software maintenance or upgrade. It is referred to here as follow-up to emphasize that the process of knowledge-base refinement that begins in the prototyping stage can continue even after the system is commissioned. The need for additional detail may come to light only after the application has been in use for some period of time. The time required to implement such enhancements should be taken into account.

Implementation Tools. In the mid to late 1980s, the expert system tools market was flooded with a plethora of expert system shells of every imaginable flavor, all varying widely in their representational and development capabilities. The 1990s have seen the market narrowing considerably; many of the tools have converged on a common set of features. Nevertheless, choosing an appropriate tool for knowledge-based application development is not easy. A number of knowledge level criteria have to be considered, including representational capabilities, eg, rules and objects; knowledge organization capabilities, eg, rule sets and taxonomies; reasoning capabilities, eg, backward and forward chaining, rule prioritization; uncertainty handling features, eg, certainty factors and belief values; truth maintenance or assumption maintenance capabilities; and special-purpose features such as temporal reasoning capabilities, which may be needed for real-time applications, or planning and scheduling systems.

In addition, other capabilities relevant to system development, end user interaction, and application integration have to be taken into account: development features, eg, graphic displays of knowledge organization, knowledge-base maintenance features, explanation facilities, system documentation capability, end user interface development tools, availability of an integrated programming language for custom coding, access to external data files,

interface to external programs, hardware platforms supported, vendor reliability and service, and price.

In selecting an appropriate tool, two other factors may also be important. First, it may be beneficial to leverage higher initial cost for a more full featured tool, if future applications are likely to need those capabilities. Second, more than one tool may be necessary to serve the needs of different types of applications, ie, the adage "one size does not fit all" holds well in the tool business. Available tools in the market can be categorized into roughly three classes.

Low End Tools. These are suitable for learning about the technology and building small applications, but have limited representational capabilities. Typically, the tools in this category are those that a domain expert, not knowledgeable in expert systems technology, might use. Low end tools are also suitable for providing hands-on training to people new to the technology. Some examples of commercially available tools in this category are VP-Expert (Paperback Software) and Exsys (Exsys Corp.).

General-Purpose Tools of Medium Complexity. These are tools that provide multiple representation schemes, and capabilities such as uncertainty handling and knowledge organization. Typically, they also have reasonable interfaces and application integration capabilities. Tools in this category are applicable in most situations that are beyond the simplistic. However, for certain applications, these tools may fall short; for example, if the application is very large, requires real-time capabilities, or extensive custom coding. Typically, medium complexity is also associated with prices that are middle-of-the-road. Two tools that exemplify this category are Nexpert Object (Neuron Data) and Kappa-PC (Intellicorp).

High End Tools and Special-Purpose Tools. These tools are suitable for those applications that have complex requirements, or those that clearly map into a particular type of problem-solving task. This category includes tools that have extensive development capabilities, or features specifically designed to be used for a particular type of task. Examples of such special-purpose features are interfaces to process control systems, temporal reasoning capabilities, or scheduling capabilities. G2 (Gensym Corp.) is a good example of a tool with some of these features. Examples of task-specific tools are TestBench (Carnegie Group) and Design π (Design Power, Inc.). ProKappa (Intellicorp) is an example of a tool with fairly sophisticated features, at the high end of the spectrum of representational capabilities.

Others. In rare cases, there may be a fourth category, ie, implementation in a programming language such as Lisp or Prolog. This approach was popular in the early days of AI, primarily because of the rapid prototyping capabilities of such languages. However, from the viewpoint of good software engineering, this is probably not an advisable direction to take. The tool approach is preferable; it has the advantages of explicit representation, built-in capabilities for documentation, explanation and integration, and if needed, access to a programming language as well.

Alternative Technologies

For a number of cognitive or interpretive tasks, there are alternatives to mainstream knowledge-based systems that may be more appropriate, especially if adaptive behavior and learning capability are important to system performance. Two approaches that embody these characteristics are neural networks (nets) and case-based reasoning.

Neural Nets. Neural nets arose out of an alternative approach to solving the problems raised by AI. Instead of modeling high level reasoning processes, neural networks attempt to produce intelligent behavior using the brain as the architectural model. The motivation behind neural network research is the modeling of cognitive activities such as perception, learning, and data interpretation, which symbolic AI approaches do not address adequately. From a practical standpoint, neural nets attempt to address many limitations of knowledge-based systems, including lack of adaptability (self-modification in response to a changing world); lack of robustness (tolerance for missing, bad, or incomplete information); problems of knowledge acquisition (extracting knowledge from experts); problems of storage capacity (amount of knowledge that can be stuffed into a knowledge base); problems of scalability (performance of large systems in comparison to small ones); and problems of speed, especially with increase in the size of the system.

Since biological systems can reasonably cope with some of these problems, the intuition behind neural nets is that computing systems based on the architecture of the brain can better emulate human cognitive behavior than systems based on symbol manipulation. Unfortunately, the processing characteristics of the brain are as yet incompletely understood. Consequently, computational systems based on brain architecture are highly simplified models of their biological analogues. To make this distinction clear, neural nets are often referred to as artificial neural networks.

Neural nets consist of many individual computing elements that are interconnected (61). Each individual computing element, or node, is computationally simple, yet complex behavior can arise out of the multiplicity of the nodes and their interconnectivity. The physical realization of this computing machinery can either be in hardware form, or in the form of a software emulation run on a conventional computer. Instead of symbolic representations, the knowledge inside a neural net is distributed all over the network in the form of connection strengths between computing elements. There is no knowledge engineering; neural nets automatically acquire the connection strengths when they are trained on representative examples of inputs and outputs. Figure 5 shows the architecture and the elemental computing capability of a typical neural net.

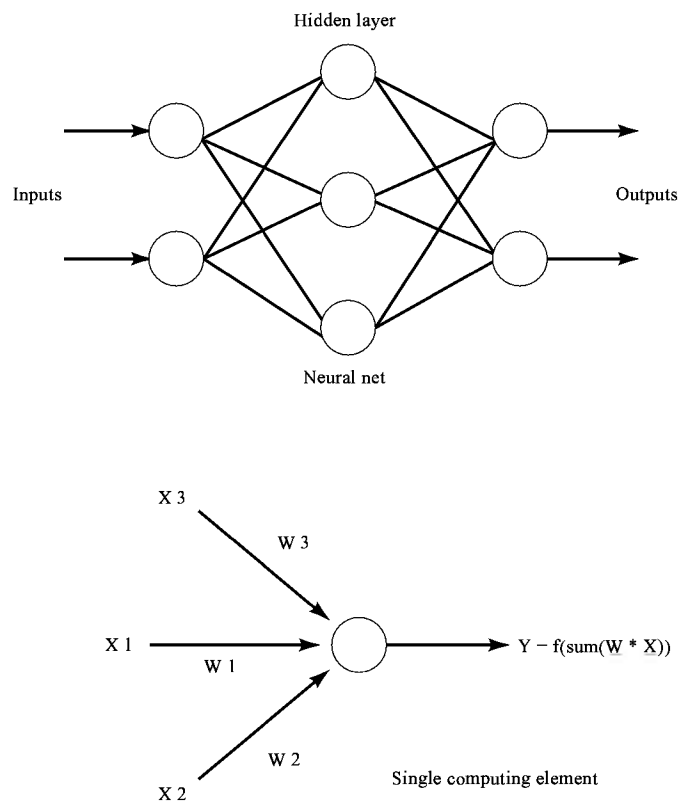


Fig. 5. The backpropagation neural net.

Neural net architectures come in many flavors, differing in the functions used in the nodes, the number of nodes and layers, their connectivity, and

most importantly, the training algorithm used to make the net learn the connection strengths. The architecture shown in Figure 5 represents one particular type of neural net known as a backpropagation network. Backpropagation networks are the most popular and well-understood types of neural nets, and they have been used in many different types of applications. Other network architectures include self-organizing feature maps, adaptive resonance theory networks, radial basis function networks, and learning vector quantization networks (62–65).

Neural networks have the following advantages: (1) once trained, their response to input data is extremely fast; (2) they are tolerant of noisy and incomplete input data; (3) they do not require knowledge engineering and can be built directly from example data; (4) they do not require either domain models or models of problem solving; and (5) they can store large amounts of information implicitly.

In general, neural nets are particularly good at pattern recognition, classification, and data interpretation tasks. This makes them especially suitable for a variety of applications: classifying input signal patterns into one or more predetermined categories, eg, identification of compressor problems based on vibration patterns; feature recognition in image processing; identifying categories in raw data, eg, organizing data into related clusters; and complementing knowledge-based systems in areas where KBS technology is notoriously weak, eg, mapping raw data into symbolic abstractions with which the KBS can then reason.

Neural nets can also be used for modeling physical systems whose behavior is poorly understood, as an alternative to nonlinear statistical techniques, eg, to develop empirical relationships between independent and dependent variables using large amounts of raw data.

When considering neural network technology as an alternative to knowledge-based systems, there are certain limitations of neural nets that should be recognized. Neural nets, in general, do not handle time well. Their design does not support end-user interaction, eg, asking the user for missing data. They do not have explanation capability. Their performance is usually highly dependent on the amount and quality of training data used. Some network architectures, eg, backpropagation, require long training times. Many design issues are more art than science, and often have to be resolved by trial and error, eg, the number of nodes and layers to use, the appropriate connectivity between nodes, and the amount of training needed.

In the chemical engineering domain, neural nets have been applied to a variety of problems. Examples include diagnosis (66,67), process modeling (68,69), process control (70,71), and data interpretation (72,73). Industrial application areas include distillation column operation (74), fluidized-bed combustion (75), petroleum refining (76), and composites manufacture (77).

Case-Based Reasoning. Case-based reasoning is a relatively new alternative to expert systems. The case-based approach is founded on the observation that humans often solve problems by drawing upon specific cases or episodes from their past experience. As a simplification, one might say that the emphasis in case-based reasoning is on memory rather than knowledge. The types of memory addressed include semantic memory (how static facts, and their relation to one another, are stored), episodic memory (how information about temporal events is stored), and conceptual memory (how concepts are stored in episodic form). By modeling memory, case-based reasoning theory attempts to cohesively explain many aspects of cognition, learning, and problem solving, bridging research in both AI and psychology.

Part of the motivation for case-based reasoning research stemmed from the same limitations of knowledge-based system technology that motivated neural nets. That is, knowledge acquisition from experts can be a notoriously difficult task; knowledge-based systems have neither memory nor evolutionary behavior, ie, they do not remember the results of their problem solving, and they do not learn to perform differently based on past successes or failures; and some knowledge-based systems do not have the ability to degrade gracefully, ie, when confronted with situations that are not explicitly covered in the knowledge base, they fail without providing partial solutions.

However, the way case-based reasoning (CBR) addresses these limitations is fundamentally different from that of neural nets. CBR systems make a fundamental commitment to representing knowledge in the form of episodic cases, which are then retrieved, adapted to new situations, and stored for future reference. The case is the basic unit of knowledge in a CBR system. The knowledge base is the case memory, and the problem solving strategy is the mechanism by which cases are retrieved, modified, applied to the new situation, and then stored back into the case repository. The key computational issues are how to appropriately index cases for retrieval, what methods to use for adapting old cases to new situations, what to do if the adapted case fails to work, and how to store new cases (successful or otherwise) in the case memory. Several good overviews of the technology are available (78,79). More extensive coverage is provided in recent texts (80).

Case-based reasoning, as a general model of problem solving, has several advantages: knowledge acquisition, although not eliminated, is simplified, eg, only actual episodic cases have to be extracted from the expert, not generalized rules; problem solving behavior can evolve through repeated application of the case memory, ie, CBR systems can learn from experience; and problem solving behavior can be robust, ie, even if retrieved cases fail to apply exactly to a new situation, they may work after modification.

At the same time, there are also many open research issues for which general solutions are not yet available, including the content and structure of cases, the organization and indexing of cases, the problems associated with search as case memories grow larger, the evolution of indexing schemes with experience, metrics for deciding which cases are applicable to a new situation, mechanisms for modifying cases, and mechanisms for repairing solutions when modified cases fail to work. In spite of these open-ended problems, a number of successful applications have been reported (81–84); however, the bulk of these applications have been in academic settings. The technology as a whole has been slow to move into the industrial world. This is partly because of the complexity of the case representation and the unavailability of easy-to-use CBR tools. This contrasts with the representational clarity and programming simplicity that were responsible for the knowledge-based system explosion.

At the present time, there are almost no well-known applications of CBR technology in the chemical engineering domain. A few industrial applications in other domains have been reported (85,86). The applications cover problem-solving tasks such as scheduling, planning, design, and diagnosis. A few case-based development tools have also emerged. These include CBR-Express (Inference Corp.) and Remind (Trinzic Corp.).

BIBLIOGRAPHY

1. W. J. Clancey, "Viewing Knowledge Bases as Qualitative Models," *IEEE Expert* (Summer 1989).

2. E. H. Shortliffe, *Computer-Based Medical Consultations: MYCIN*, Elsevier, New York, 1976.

3. J. McDermott, *Art. Intell.* **19** (1982).

4. R. O. Duda, J. G. Gaschnig, and P. E. Hart, in D. Michie, ed., *Expert Systems in the Microelectronic Age*, Edinburgh University Press, UK, 1979.

5. A. M. Turing, *Mind* **59** (1950).

6. J. Haugeland, *Artificial Intelligence: The Very Idea*, MIT Press, Boston, 1989.

7. C. E. Shannon, *Sci. Am.* **182**, 48 (1950).

8. A. Newell, H. A. Simon, and J. C. Shaw, in E. A. Feigenbaum and J. Feldman, eds., *Computers and Thought*, McGraw-Hill Book Co., Inc., New York, 1963.

9. G. W. Ernst and A. Newell, *GPS: A Case Study in Generality and Problem Solving*, Academic Press, Inc., New York, 1969.

10. R. C. Schank, *Conceptual Information Processing*, North-Holland, Amsterdam, 1975.

11. T. Winograd, *Understanding Natural Language*, Academic Press, Inc., New York, 1972.

12. M. Minsky, *Semantic Information Processing*, MIT Press, Cambridge, Mass., 1968.

13. E. Rich, *Artificial Intelligence*, McGraw-Hill Book Co., Inc., New York, 1983.

14. R. E. Fikes and N. J. Nilsson, *Artif. Intell.* **2** (1971).

15. R. K. Lindsay and co-workers, *Applications of Artificial Intelligence to Chemistry: The DENDRAL Project*, McGraw-Hill Book Co., Inc., New York, 1980.

16. D. L. Chester, D. E. Lamb, and P. Dhurjati, *ASME-Computers in Engineering* **1**, 345–351 (1984).

17. J. Calandranis, G. Stephanopoulos, and S. Nunokawa, *Chem. Eng. Prog.* **86** (1990).

18. T. S. Ramesh, J. F. Davis, and G. M. Schwenzer, *Computers Chem. Eng.* **16**(2), (1992).

19. C. A. Siletti and G. Stephanopoulos, *A Computer Program for Synthesizing Downstream Process Designs*, technical report no. LISPE-88-043, Laboratory

- for Intelligent Systems in Process Engineering, MIT, Cambridge, Mass., 1988.
20. D. R. Myers, J. F. Davis, and D. J. Herman, *Computers Chem. Eng.* **12**(9/10) (1988).
21. S. H. Rich, A. L. Guerrero, and G. M. Schwenzer, "Knowledge-based Simulation for Operations Analysis and Facilities Design of Refinery Off-Sites," paper presented at *NPRA Computer Conference*, Seattle, Wash., 1990.
22. T. Weatherill and I. T. Cameron, *Computers Chem. Eng.* **13**(11/12) (1989).
23. V. K. Tzounas and co-workers, *Ind. Eng. Chem. Res.* **29**(3) (1990).
24. R. Lakshmanan and G. Stephanopoulous, *Computers Chem. Eng.* **12**(9/10) (1988).
25. M. Realff, *An Analysis of the Chemical Batch Production Problem and a Detailed Methodology for Scheduling*, technical report no. LISPE-88-054, Laboratory for Intelligent Systems in Process Engineering, MIT, Cambridge, Mass., 1989.
26. M. Hofmeister, L. Halasz, and D. W. T. Rippin, *Computers Chem. Eng.* **13**(11/12), 1255–1261 (1989).
27. J. F. Faccenda and T. E. Baker, "An Integrated AI/OR Approach to Blocked Operations Scheduling," paper presented at the *NPRA Computer Conference*, Seattle, Wash., 1990.
28. B. Chandrasekaran, *IEEE Expert* **1**(3), 23–30 (1986).
29. A. Bundy, *The Computer Modeling of Mathematical Reasoning*, Academic Press, Inc., London, 1983.
30. K. Parsaye and K. Y. Lin, "An Expert System Structure for Automatic Fault Tree Generation for Emergency Feedwater Systems for Nuclear Power Plants," in *Proceedings of the Second IEEE Westex Conference*, Anaheim, Calif., June, 1987.
31. P. Jackson, *Introduction to Expert Systems*, Addison-Wesley, Worthington, UK, 1986.
32. L. A. Zadeh, *Inf. Control* **8**, 338–353 (1965).
33. D. McDermott, *Cognitive Sci.* **6**, 101–155 (1982).
34. D. McDermott and J. Doyle, *Artif. Intell.* **13**(1) (Nov. 1980).
35. T. C. Bylander and S. Mittal, *AI Magazine* **7**, 66–77 (1986).
36. K. Parsaye and M. Chignell, *Expert Systems for Experts*, John Wiley & Sons, Inc., New York, 1988.
37. W. J. Clancey, *Artif. Intell.* **20** (1983).
38. M. C. Tanner and J. Josephson, "Explanation and Abductive Justification," in *Proceedings of the Spring Symposium Series on Artificial Intelligence in Medicine*, Stanford University, Calif., Mar. 1988.
39. W. J. Clancey in J. F. Traub, ed., *Annual Reviews*, Palo Alto, Calif., 1986.
40. B. Chandrasekaran and S. Mittal, *Intl. J. of Man-Machine Studies* **19**, 425–436 (1983).
41. T. Bylander, "Complexity of Model-Based Reasoning" in *Proceedings of the 1989 AAAI Workshop on Model-Based Reasoning*, Seattle, Wash., 1989.
42. J. McDowell and J. F. Davis, *AIChE J.* **37**(4) (Apr. 1991).
43. V. Venkatasubramanian and S. H. Rich, *Computers Chem. Eng.* **12**(9/10), 903–921 (1988).
44. R. Davis and W. Hamscher, in H. E. Shrobe, ed., *Explorations in Artificial Intelligence*, MIT Press, Cambridge, Mass., 1988.
45. D. Dvorak and B. Kuipers, *IEEE Expert* **6**(3) (June 1991).
46. L. Steels, *AI Magazine* **11**(2) (Summer 1990).
47. R. T. Dodhiawala, N. S. Sridharan, and C. Pickering, in V. Jaganathan, R. T. Dodhiawala and L. Baum, eds., *Blackboard Architecture and Applications*, Academic Press, Inc., New York, 1989.
48. T. J. Laffey and co-workers, *AI Magazine* **9**(1) (Spring 1988).
49. R. Bhatnagar, B. Chandrasekaran, and D. D. Sharma, *Communications of the ACM* **34**(8) (1991).
50. P. A. Sachs, A. M. Patterson, and M. H. M. Turner, *Expert Systems* **3**(1) (1986).
51. S. Padalkar and co-workers, *IEEE Expert* **6**(3) (June 1991).
52. J. E. Clancy, G. J. Carrette, and K. M. Gersten, *ISA Transactions* **31**(3) (1992).
53. S. Weiss and C. Kulikowski, *A Practical Guide to Designing Expert Systems*, Rowman and Allenheld, Totowa, N.J., 1984.
54. E. Payne and R. McArthur, *Developing Expert Systems*, John Wiley & Sons, Inc., New York, 1990.
55. E. Sacerdoti, *AI Expert*, (May 1991).
56. J. R. Slagle and M. R. Wick, *AI Expert*, (Winter 1988).
57. B. Chandrasekaran, *AI Magazine* **11**(4) (Winter 1990).
58. S. K. Shum and co-workers, "A Knowledge-Based System Framework for Diagnosis in Process Plants," in *Proceedings of the Seventh Power Plant Dynamics, Control and Testing Symposium*, Knoxville, Tenn., 1989.
59. J. S. di Piazza and F. A. Helsabeck, *AI Magazine* **11**(3) (1990).
60. T. A. Nguyen and co-workers, *AI Magazine* **8**(2) (1987).
61. R. Lippmann, *IEEE ASSP Magazine* **4** (Apr. 1987).
62. T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, 1987.
63. G. A. Carpenter and S. Grossberg, *Appl. Opt.* **26**, 4919–4931 (1987).
64. J. Moody and C. J. Darken, *Neural Computation* **1**, 281–294 (1989).
65. S. C. Ahalt and co-workers, *Neural Networks* **3**, 277–290 (1990).
66. J. C. Hoskins, K. M. Kaliyur, and D. M. Himmelblau, "The Application of Artificial Neural Networks to Fault Diagnosis in Chemical Processing," paper presented at *AIChE Spring Meeting*, Houston, Tex., 1988.
67. J. J. Ferrada, M. D. Gordon, and I. W. Osborne-Lee, "Application of Neural Networks for Fault Diagnosis in Plant Production," paper presented at *AIChE National Meeting*, San Francisco, 1989.
68. J. C. Hoskins and D. M. Himmelblau, *Computers Chem. Eng.* **12**, 881–890 (1988).
69. N. Bhat and T. J. McAvoy, "Dynamic Process Modeling via Neural Computing," paper presented at *AIChE National Meeting*, San Francisco, 1989.
70. N. Bhat and T. J. McAvoy, *Computers Chem. Eng.* **14**, 573–583 (1990).
71. L. H. Ungar, B. A. Powell, and S. N. Kamens, *Computers Chem. Eng.* **14**, 561–572 (1990).
72. J. R. Whiteley and J. F. Davis, "Qualitative Interpretation of Sensor Patterns using a Similarity-Based Approach," paper presented at the *IFAC Symposium on On-Line Fault Detection and Supervision in the Chemical Process Industries*, Newark, Del., Apr. 1992.
73. T. J. McAvoy and co-workers, "Interpreting Biosensor Data via Backpropagation," in *Proceedings of the International Joint Conf. on Neural Networks*, Washington, D.C., 1989.
74. D. Blanchard, *AI Week*, (Sept. 1, 1990).
75. S. Moorthy and P. Iyer, "Neural Network Simulation for Plant Control Optimization," in *Proceedings of the Industrial Computing Conference*, ISA, Anaheim, Calif., 1991.
76. G. Hobson, "Neural Network Applications at PSP," paper presented at *NPRA Computer Conference*, Seattle, Wash., 1990.
77. D. A. Sofge and D. A. White, "Neural Network Based Process Optimization and Control," in *Proceedings of the 29th Conference on Decision and Control*, 1990.
78. S. Slade, *AI Magazine* **12**(1) (1991).
79. J. L. Kolodner, *AI Magazine* **12**(2) (1991).
80. C. Riesbeck and R. C. Schank, *Inside Case-Based Reasoning*, Lawrence Erlbaum, Hillsdale, N.J., 1989.
81. E. Bareiss, B. Porter, and C. Wier, *J. of Man-Machine Studies* **29**, 549–561 (1988).
82. K. Sycara, "Using Case-Based Reasoning for Plan Adaptation and Repair," in *Proceedings of the DARPA Workshop on Case-Based Reasoning*, Morgan

Kaufman, San Mateo, Calif., 1988.

83. D. Navinchandra, "Case-Based Reasoning in CYCLOPS, A Design Problem Solver," in Ref. 82.

84. R. Turner, "Organizing and Using Schematic Knowledge for Medical Diagnosis," in Ref. 82.

85. P. Koton, "SMARTPLAN: A Case-Based Resource Allocation and Scheduling System," in *Proceedings of a Workshop on Case-Based Reasoning*, Morgan Kaufman, San Mateo, Calif., 1989.

86. W. Mark, "Case-Based Reasoning for Autoclave Management," in Ref. 85.

General References

Artificial intelligence J. Haugeland, *Artificial Intelligence: The Very Idea*, MIT Press, Boston, Mass., 1989.

E. Rich, *Artificial Intelligence*, McGraw-Hill Book Co., Inc., New York, 1983.

P. H. Winston, *Artificial Intelligence*, Addison-Wesley, Reading, Mass., 1984.

Expert systems

J. P. Ignizio, *Introduction to Expert Systems*, McGraw-Hill Book Co., Inc., New York, 1991.

P. Jackson, *Introduction to Expert Systems*, Addison-Wesley, Worthington, UK, 1986.

K. Parsaye and M. Chignell, *Expert Systems for Experts*, John Wiley & Sons, Inc., New York, 1988.

M. Van Horn, (The Waite Group), *Understanding Expert Systems*, Bantam Books, New York, 1986.

D. A. Waterman, *A Guide to Expert Systems*, Addison-Wesley, Reading, Mass., 1986.

Expert system development

P. Harmon and R. Maus, *Expert Systems Tools and Applications*, John Wiley & Sons, Inc., New York, 1989.

E. Payne and R. McArthur, *Developing Expert Systems*, John Wiley & Sons, Inc., New York, 1990.

S. Weiss and C. Kulikowski, *A Practical Guide to Designing Expert Systems*, Rowman and Allenheld, Totowa, N.J., 1984.

Expert systems in chemical engineering

R. H. Cadmus and M. D. Woolsey, *Oil & Gas J.*, (Jan. 9, 1989).

Chem. Eng. Prog., Expert Systems Issue, **83**(9) (1987).

M. Mavrovouniotis, ed., *Artificial Intelligence Applications in Process Engineering*, Academic Press, Inc., New York, 1990.

D. A. Rowan, *AI Expert*, (Aug. 1989).

G. Stephanopoulous, *Computers Chem. Eng.* **14**(11), (Nov. 1990).

T. S. Ramesh
Mobil Research and Development Corporation

EXPLOSIVELY CLAD METALS.

See METALLIC COATINGS.